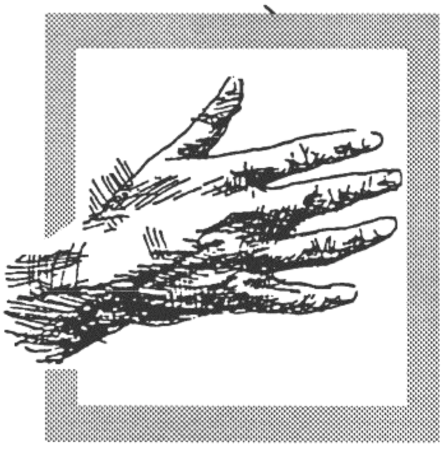
FORSCHUNGSVERBUND

ANWENDUNGEN DER KI IN NRW

Virtual Environments and Situated Agents

Ipke Wachsmuth / Yong Cao

Arbeitsgruppe Wissensbasierte Systeme
(Künstliche Intelligenz)
Technische Fakultät
Universität Bielefeld



PROJEKT VIENA

KI-NRW 93 - 11

Virtual Environments and Situated Agents

Ipke Wachsmuth and Yong Cao
Faculty of Technology, University of Bielefeld,
P.O. Box 100131, 33501 Bielefeld, Germany

1 Introduction

Virtual environments are highly interactive means for experiencing and manipulating three-dimensional scenes. Along with sophisticated techniques for natural visualization and rapidly increasing computer power of modern graphics workstations, virtual environments are becoming highly attractive for design and simulation. One area in which this new media might prove especially useful is architecture and, in particular, interior design. For example, the visualization of an office room or a building prior to its physical realization could help a designer to obtain realistic impressions of a construction while it is evolving and to give free way to imagination at the same time. It is one of the aims, eventually, that a designer could be able to explore, and interact with, a manipulable environment without wasting physical matter and with the ability to readily change the immaterial model.

While progress has still to be made before virtual environments can be used other than with largely precomputed models, the issue of interactive modeling has its own difficulties and challenges. A comfortable human-computer interface seems important which can keep the designer free from technical considerations such as planning of geometric detail. Some researchers have begun using the data glove for rearranging objects in a scene (Böhm et al. 1992). To a human it seems more natural to grasp a chair, lift it from the ground, and put it down at a new position than calculating an exact target position for changing the geometric model. As simulation of physical laws in virtual environments is progressing, such direct manipulation has become most attractive for virtual design. A natural interaction with a virtual environment seems impaired, though, as long as a designer cannot use *language* and *symbols* as a means of communication in virtual design.

In this paper a scenario for one of the projects in a new research program on "Artificial Intelligence and Computer Graphics" at the University of Bielefeld is presented. In the VIENA project ("Virtual Environments and Agents") we want to provide ways of intelligent communication with a technical system for designing and generating 3D computer graphics[1]. To do so we are going to apply new AI

[1]Research in the VIENA Project is partly supported by the Ministry of Science and Research of the federal state North-Rhine-Westphalia under grant no. IVA3-107 007 93. The ideas and conclusions expressed are those of the authors and do not necessarily reflect the views of the funding institution. The authors are indebted to Norbert Siekmann, Britta Lenzmann, and Majid Amanzadeh for assisting during the research.

methods and techniques that build on ideas of situated communication and agents. We use the notion of agent to refer to an overall system which integrates aspects of perception, action, and communication and which is able to use these faculties with respect to a given task and in a given situation. We think of an agent as a delegate "who" cooperates with a designer in an overlapping perceptual situation and serves as an "intelligent mediator." An agent will communicate with a user during performance of a task and will help the user exploit internal scene information not readily available. An agent could also help to manipulate a scene according to the expectations of the user where it is difficult to use direct manipulation or where physical laws are not in effect.

To master the communication with the user, the "intelligent mediator" is informed about the actual scene as it is seen from the perspective of the user. As one feature a synthetic agent graphically visualized could be used to place the designer's eye in the virtual environment so the designer is situated in the developing scene. If a designer could use language input from the perspective of his or her virtual presence, the interaction modalities would be greatly enhanced.

In the following section, we discuss further ideas about virtual environments, their possible contribution to architectural design, and some research questions for virtual environments. Among the most urgent research questions an enhancement of software techniques for interactive modeling has been recognized. In Section 3, we sketch different models and notions of agents that are currently found in literature. We also give some ideas on what kind of jobs could be extended to an agent that mediates between a user and the technical system. In the fourth section, we give more details about our scenario and particularly deal with the topic of situatedness. The fourth section also includes some remarks on spatial reasoning which is a key feature for an agent that serves to mediate a designer's interaction in a virtual 3D-environment.

2 Virtual Environments

The use of computer graphics for accurate and natural visualization has become increasingly important in all areas of object and scene design. Not far ago, the output of conventional graphics systems was exclusively thought for viewing by a human user, that is, there was little possibility to interact with the displayed images. A major goal of current research and development is to bridge the gap between high-quality visualization systems which present output to a passive user, and interactive systems which are able to accept and display user interventions in a visualized scene as soon as they are issued. In the ideal case, the user is immersed in a scene directly, with all senses, and is able to interact with objects in the scene. This is the idea of a virtual environment.[2]

[2]Merriam Webster's New Collegiate Dictionary, Ninth Edition, defines "virtual" as "being in effect but not in actual fact", and "environment" as "the conditions, circumstances, and influences surrounding and effecting an organism."

In accordance with the executive summary of research directions in Virtual Environments resulting from an NSF Invitational Workshop (cf. Bishop et al. 1992), we refer to "virtual environments" as real-time interactive graphics with three-dimensional models, when combined with a display technology that gives the user immersion in the model world and direct manipulation. A virtual environment, thus, is not simply an improvement of conventional display techniques but it is a new media of communication with novel kinds of challenges for computer and communication technology. To us, a virtual environment does not necessarily involve use of a head phone or a data glove; there are other ideas how a responsive environment can give a user the feeling of being immersed. For example, a positional tracking system like the Polhemus (Raab et al. 1979) could be used for virtual pointing, and objects pointed to could be addressed by a verbal command.

While the initial development of virtual environment technology was pushed through military and space research and its commercial use by entertainment industry, a wide field of application is now developing in science and technology, medicine, art, and architecture (Krüger 1993). Examples for the use of virtual environments in these areas are interactive exploration models like the "virtual wind tunnel" of NASA Ames (Bryson and Levit 1992) and the "walkthrough environment" of the University of North Carolina with applications in architecture and medicine (Brooks 1987). New applications are also discussed in the simulation and visualization of telepresence and robotics (e.g., Kirsch et al. 1993), and in the design disciplines as urban planning and architectural modeling of buildings and building interiors.

The use of virtual environments in design, which is our focus, could contribute to the following goals:

- three-dimensional pre-exploration of geographical sites, building construction-plans, interior architecture
- interactive design of scene models, scene objects, illumination, surface textures, and object interplay with given surroundings like a municipal area
- a test of sensual experiences like proportion, coloring, light-and-shadow, surface structures, e.g., by simulating the three-dimensional experience of a synthetic building interior with respect to design and illumination aspects.

An example for the usage of virtual environments in interior design is the three-dimensional modeling and visualization of office rooms or buildings prior to their physical realization. The aim is to obtain realistic impressions of a construction while it is evolving and to give free way to the creative process at the same time.

In this context, modeling refers to the data structures that are used to record the geometrical information for the environment. This information includes the shape of the objects, their parts and physical properties, and the ways how they interact with other objects in the environment and with the user. To give an idea, Fig. 1 shows a geometry (wireframe) model of an interior which can be used for exploring materials and illumination. In creating such a geometry model, a designer will need to communicate ideas of complex form to a technical device, and may face crucial obstacles in the process of designing. Hence, a comfortable user interface is important to keep the designer free from technical considerations such as planning of geometric details, measured proportions, etc.

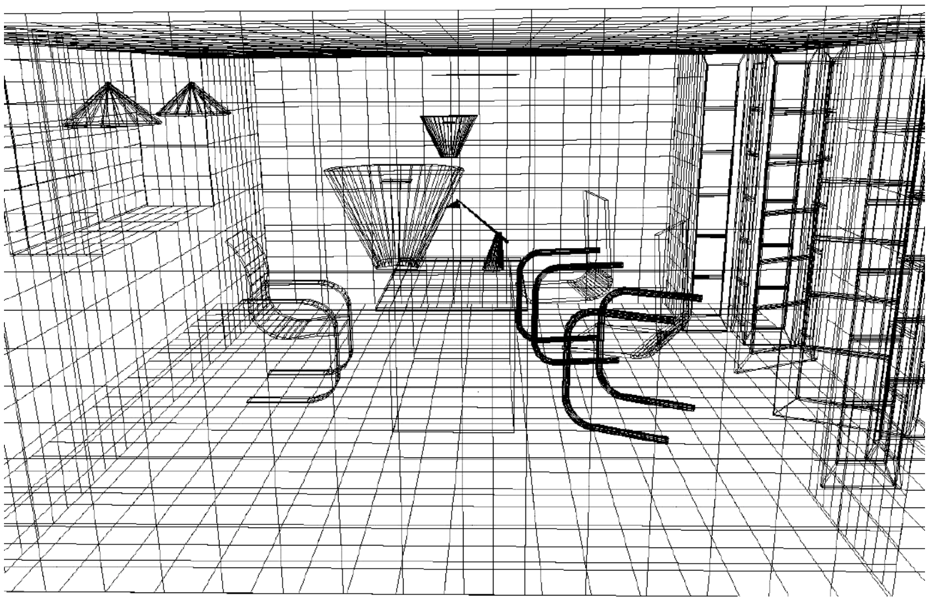


Fig. 1. Wireframe model of an object scene from interior design[3]

Among research goals for virtual environments, *modeling* has been recognized as the key issue for the present time. A list of the most pressing needs (cf. Bishop et al. 1992) includes the development of software for design of and interaction with virtual worlds that is modular, flexible, and abstract. The user must be able to move through the model, interact with objects in the model, and change the model interactively. At least part of physics should be in effect in virtual environments. For instance, when an object is dropped, it should fall until it reaches the ground, and it should not be possible for objects to pass through each other.

We think that usage of the so-far available means of interaction, e.g., the data glove, is but one way to manipulate the arrangement of a scene. With the data glove, a scene object could be grasped, moved to and placed at a different position. The advantages of such direct handling of an object – at least as it pertains to spatial manipulation – need not be mentioned. Until recently, there was no direct feed-back to control success of a desired action other than by eye inspection, for instance, to justify that a chair has actually been placed on the ground. The use of bounding plane constraints in modeling has given rise to some progress in that respect. But changing the size or material, e.g., the color of an object, would most likely involve a mental detour, for example, a "space menu" might have to be used.

An alternative way we want to explore in the VIENA project is the usage of verbally communicated instructions which are put in effect by a mediating system. That is, we want to have the choice to either change a scene by way of hand access,

[3]Fig.1 is included by kind permission of Steve Drucker, MIT Media Lab.

or to instruct the system to carry out according actions where direct manipulation seems impossible or unnatural. Eventually, we think of the use of voice input and gestures as a parallel input modality. In either case the technical execution of the scene manipulation would rely on an object-oriented geometry model of the complex scene. Extending current means of virtual environment manipulation by means of verbal instruction would add a powerful dimension since it would involve the use of symbolic communication for interacting with a virtual environment.

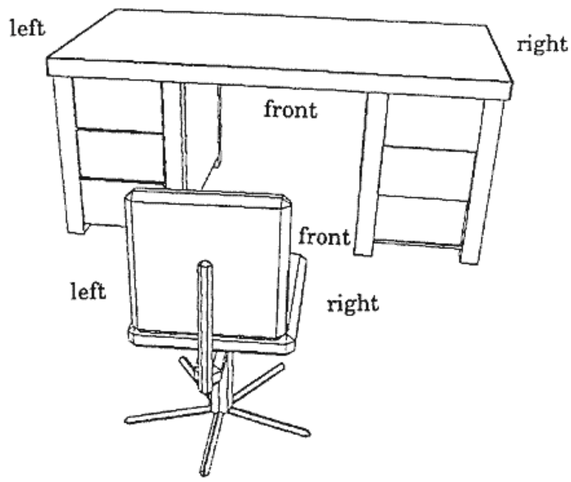


Fig. 2. Different use of particles 'left', 'front' and 'right' in a scene with a chair and a desk (adapted from Levelt 1986)

If we use verbal communication, we need to be aware of the fact that the way we refer to details in a scene is "situated." For instance, it may depend on the objects themselves where we would speak of "front" and "back," e.g. a chair and a desk impose local reference structures on space, and they may have opposite front parts (Fig. 2). We may use still different notions of "front" and "back" when making reference to our bodily presence. Hence, natural verbal communication in a virtual environment cannot be put into effect without the system knowing something about the cognitive structure of space as it is perceived by a virtual human participant.

We want to use "agents" to conceptualize and realize the mentioned mediating system. Before we explain how we go about doing so, we briefly sketch different agent notions found in the current literature, and how agent ideas have been included in example areas of research and development.

3 Agents

In the mentioned research agenda for virtual environments (Bishop et al. 1992), one theme is the development of knowledge-based agents for human-computer interfaces. Ideas of a general intelligent agent that uses knowledge to perform actions in the service of goals have been researched in AI since more than twenty years (cf. Newell 1981 for a discussion of this subject). A different perspective has been taken by Minsky (1986) in explaining intelligence as a combination of many simple processes he refers to as agents. Agents that work together can perform a job without each agent knowing anything about the job; in total intelligent behavior derives. Recent attempts to develop larger and more complex knowledge-based systems have revealed shortcomings of centralized, single-agent architectures and have acted as a springboard for research in Distributed AI (DAI; cf. Adler et al. 1992).

Different models and notions of agents are currently found in diverse literature in the fields of autonomous and distributed systems (cf. Meinkoehn and Knoll 1993), language-action systems (Chapman 1991), and in graphics animation. Of interest, in our context, is work on "interface agents" which assist users in a computer application. Such an idea has been adopted in visions of future user interfaces, for instance, in the Japanese Friend21 project (cf. Marcus 1993). In the Newton project of Apple (cf. Marcus 1993), agent programs are envisioned which are associated with metaphors for user interaction and which might include displaying "graphics agents" as human-like beings. On the other hand, work from DAI is relevant which incorporates approaches on "cooperating agents" that work together in achieving an overall task. Finally, the idea of "situated agents" which can gain and exploit information from an actual situation is of great interest to us.

Interface agents: Interface agents are semi-intelligent, semi-autonomous systems which support a user in dealing with a computer application (Laurel 1990). Typically, interface agents are personal assistants which use knowledge about tasks, habits, preferences of their users to perform actions on their behalf. Kozierok and Maes (1993) at MIT Media Lab discuss a gradual delegation of tasks to an agent as a trust relationship is built. In an example application, they want to use machine learning techniques for having agents obtain the necessary knowledge to assist their users in the task of scheduling meetings.

Graphics agents: The work of Badler et al. (1991) is aimed at the simulation of visualized synthetic agents which are subject to human constraints and restrictions ("animated Jack and Jane"). To become moving as a human person, agents are instructed step-by-step in natural language ("animations from instructions"), and their complex motion is executable in the virtual environment. Badler et al. also deal with the question how complex action commands can be decomposed in sequences of progressively primitive actions by way of task knowledge of an intelligent agent.

Cooperating agents: Multi-agent systems as discussed in DAI (Adler et al. 1992) emphasize the aspect of task-related cooperation of independent ("autonomous") agents. Agents are ascribed a basic functionality (they can solve certain problems they are given), a cooperative head (for participating in a cooperation with other agents), and communication abilities (by way of access to communication channels to other agents). An open spectrum of agent types has been considered (Müller and Siekmann 1991), reaching from primitive (sensor-driven, reactive) agents through "social agents" with a "conscious" ability of interaction. Higher agents can have knowledge of other agents and their skills. No agent has a global view on the total problem to be solved, that is, there is no central control.

Situated agents: New lines of AI research have given notice to the fact that the actions of an intelligent agent may decisively depend on its involvement in an actual situation. A situated agent integrates aspects of perception, action, and communication in one system in order to succeed in a situation without having a complete model of it (Brooks 1991). The term "situatedness" refers to the ability of an intelligent system to exploit the actual situation, to the extent possible, as a source of information in perceiving and manipulating its environment and communicating with cooperating partners.[4]

In our work, we conceive a situated agent as an "intelligent mediator." Such an agent – which may consist of many subagents – communicates and cooperates with a human user in an overlapping perceptual situation, that is, human user and situated agent have a shared virtual world. The agent's task (or the agents' task, resp.) is to support the user's actions in the virtual environment with respect to our application, the design and manipulation of a virtual environment. The agent is embedded in the technical system and integrates skills of situated scene (model) perception, action, and communication, in order to achieve an adequate system behavior. Besides using the notion of agent as a helpful metaphor for human-computer-interaction, we think that it is also useful for conceiving the architecture of the interface software.

The situated agent could mediate usage of the following aspects:

- the shape, location, dimension and motion attributes of particular scene objects
- views of scene elements from changing spatial positions
- names the designer gives to primitive objects, aggregates, or positions
- the way how particular actions can be planned and achieved in the scene shown
- resolving reference for object descriptions, e.g., "the chair by the desk"
- dealing with objects deictically referenced by pointing gestures: "this chair"
- computing situated references like "left of", "behind", "the other one"
- detecting and preventing constraint violations
- the position of currently invisible (hidden) objects: "show..."

etc.

Different subtasks could be assigned among several cooperating agents. This way, the development of the system could be attacked in a modular fashion in which the single agent has a partial responsibility.

4 Virtual Environments and Agents: The VIENA Project

The VIENA Project is one of four projects started in 1993 in the new research field of "Artificial Intelligence and Computer Graphics" at the University of Bielefeld. Our overall goal is to enable and establish an intelligent communication with a

---

[4]Situated communication is a focus theme in a newly established special research unit at the University of Bielefeld (SFB 360, "Situated Artificial Communicators").

technical device for the interactive design and exploration of 3D computer graphics. Our application area initially chosen is interior design. One specific goal is the interactive manipulation of scene objects in a virtual environment by way of language communication. A more far-reaching goal is the inclusion of speech and gestures and of user instructions that are incomplete or vague. By incorporating AI methods, in particular, agents and symbolic communication, we expect that significant improvements can be reached for the use of virtual environments.

A major focus is the construction of situated agents which, on the one hand, communicate with the human user (designer) to receive instructions, and which cooperate with each other to realize the user's instructions, for example, to change the arrangement of scene objects in the virtual environment. In Fig. 3, a schematic view is given to explain the relation between human (as instructor), agents (for mediation), and a 3D graphics system to construct and update the visualization of the current scene.

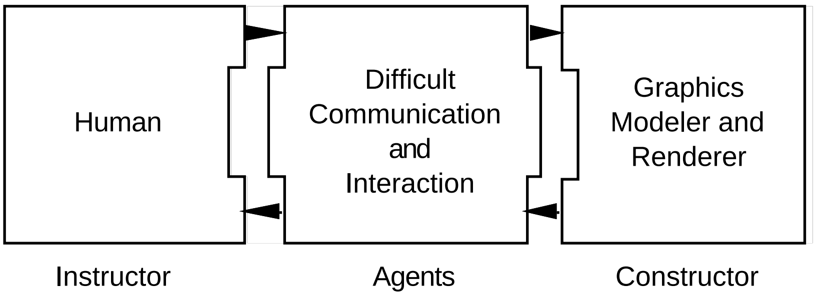


Fig. 3. Agents intelligently mediate interaction between human and virtual environment.

Our specific scenario is as follows: In a modeling session the designer (as instructor) keeps track of the evolving design (e.g., of an interior) by viewing it in a 3D display setting. The designer can change the model by communicating with the system in simple language instructions (e.g., "move the table more to the front"). The system offers a view of a resulting scene where "more" is interpreted on the basis of a relative default value. The offer can be changed by a further interaction ("still more", "not that far"), that is, the user can negotiate the semantics of an instruction.

When using verbal interaction, the system is to know about the spatial structure as perceived and experienced by the human user. For instance, the three instructions, "move the table here, more to the front, more to the shelf" incorporate different frames of spatial reference. In the first case, reference refers to the speaker's position while it is anchored externally in the second and the third case. The metrical structure does not only depend on geometry but also, for example, on how far one could reach from a position.

Thus the system – as the human – has first to find out which frame of reference is relevant, secondly, in which direction the chair is to be moved, third, how far. It should not be possible to move an object further than permitted by a physical boundary. When the table is moved, the things on the table should move with it (Fig. 3), so some agent process should take account for that. When a deictic reference is involved ("here"), the point of anchoring a reference frame needs to be clear which could be achieved by inclusion of a visualized "self" (Fig. 4).

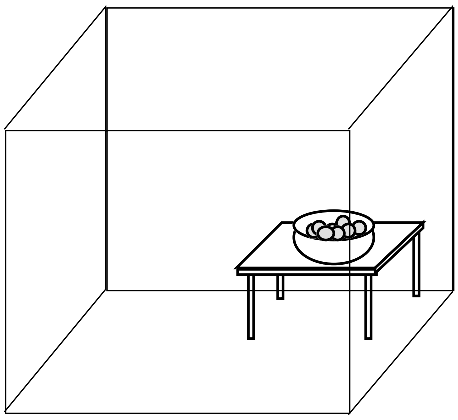


Fig. 3. "Move the table to the front!" (What happens to the bowl?) "Move the table here!" (Where is "here"?)

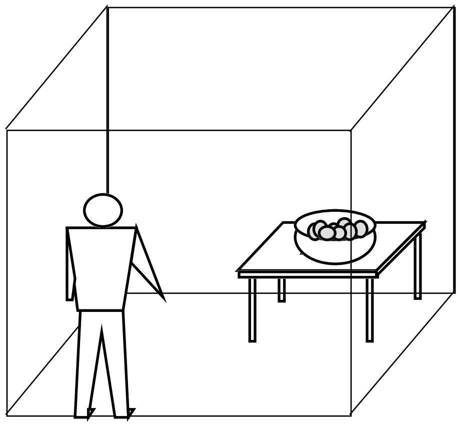


Fig. 4. "Move the table here!" (Deictic reference includes speaker's position)

Even more, the introduction of anthropomorphic features by a virtual embodied presence of the user could help to resolve ambiguity in instructions. E.g., "put the chair in front of the table" could be interpreted in different ways depending on the speaker's position (Fig. 5). Through a projection of the human asymmetries into the virtual environment, human perspectives could be brought to bear, for example, the use of situated references like "front-back" and "left-right" would be an important asset for interaction in a virtual environment.

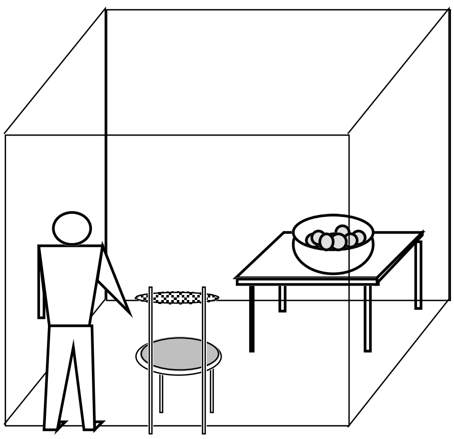


Fig. 5. "Put the chair in front of the table!" (Anthropomorphization by way of virtual "embodied" presence through situated agent)

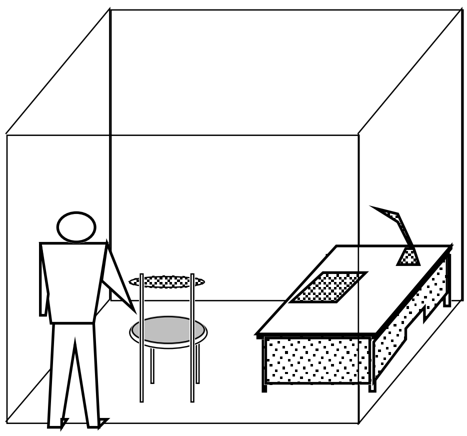


Fig. 6. "Put the chair in front of the desk!" (Modification of space through objects present)

Finally, the virtual environment is modified by the objects present in that an object could impose its own reference scheme. For example, the phrase term "in front of" is most likely interpreted in different ways – by a human, and should be so by the system – in the scenes shown in Fig. 5 and 6.

We want to exploit theories and findings from cognitive research to deal with these topics. For example, object schemes have been proposed (Lang 1989) that describe the way space is modified by the objects present. The visual-perceptive system of a human gives rise to spatial relations between object and the observer's body (Bryant 1991, 1992). Different frames of reference have been proposed: (1) the egocentric frame, defined by the three body axes: head/foot, front/back, and left right where only left/right is biologically symmetric; (2) the allocentric frame, defined by orthogonal axes external from the observer. Such axes can be anchored in a prominent landmark in the environment or be oriented according to global directions (Cao 1993). Head/foot, in gravity, is identified with top/down as long as the observer is in upright position. These axes are experienced differently in other environments (e.g., zero gravity in outer space; cf. Friederici 1989).

We conceive agents as systems with restricted ability they bring to bear with respect to a given instruction. A major obstacle in scene modeling is the technical realization of situated instructions in an exact geometry model. Thus, some agents are to carry out spatial inferences according to the expectations of a human participant. In doing so they are to exploit the current situation to the maximum extent possible. For example, by using information about most previous manipulations, possible ambiguity in an instruction could be resolved. Other agents know about current locations and materials of objects, still others about how to change a color or an illumination ("darker!"), etc. The more competent, by agent mediation, the system becomes, the more successful the designer's instructions can be interpreted and executed.

We have installed a Silicon Graphics Indigo ELAN R4000 for the main demonstrator on this project as it supports the real-time hardware shading we make use of. We also have the stereo option, for better 3D impression, which we intend to use later. We do not include real-time texturing which is not supported by this machine. Currently, we use SOFTIMAGE for scene modeling and rendering. Besides this, several Sun SPARC stations are available for algorithm and model development. These could also be the site for some of the agents, as we intend to make use of interprocess communication.

References

Adler, M., Durfee, E., Huhns, M., Punch, W., and Simoudis, E. (1992). AAAI workshop on cooperation among heterogeneous intelligent agents. *AI magazine 13*(2), 39-42.

Badler, N.I., Webber, B.L., Kalita, J. and Esakov, J. (1991). Animations from Instructions. In N.I. Badler, B.A. Barsky and D. Zeltzer (eds.): *Making Them Move - Mechanics, Control, and Animation of Articulated Figures* (pp. 51-93). San Mateo (Cal.): Morgan Kaufmann.

Bishop, G., Fuchs, H. et al. (1992). Research Directions in Virtual Environments: Report of an NSF Invitational Workshop (March 23-24, 1992, University of North Carolina at Chapel Hill). *Computer Graphics 26*(3), 153-177.

Böhm, K., Hübner, W. and Väänänen, K. (1992). GIVEN: Gesture driven interactions in virtual environments: A toolkit approach to 3D interactions. *Proc. Conf. Interface to Real and Virtual Worlds* (Montpellier, France, March 1992).

Brooks, F. (1987). Walkthrough – A Dynamic Graphics System for Simulating Virtual Buildings. *Proc. of ACM Workshop on 3D Graphics* (Chapel Hill, Oct. 1987), 9-21.

Brooks, R.A. (1991). Intelligence without reason. *Proceedings IJCAI-91*, 569-595.

Bryant, D.J. (1991). *Perceptual Characteristics of Mental Spatial Models* (Dissertation). Stanford University: Dept. of Psychology.

Bryant, D.J. (1992). A Spatial Representation System in Humans. *PSYCOLOQUY 3*(16) space.1

Bryson, S. and Levit, C. (1992). The Virtual Wind Tunnel. *IEEE Computer Graphics and Applications 12*(4), 25-34.

Cao, Y. (1993). *Zur Darstellung und Verarbeitung von Wissen über Himmelsrichtungen – Geometrische und Kognitionswissenschaftliche Aspekte* (Dissertation, University of Hamburg Dept. of Computer Science). Sankt Augustin: Infix (DISKI-33).

Chapman, D. (1991). *Vision, Instruction, and Action.* Cambridge (MA): MIT Press.

Friederici, A.D. (1989). Raumreferenz unter extremen perzeptuellen Bedingungen: Perzeption, Repräsentation und Sprachliche Abbildung. In Ch. Habel, M. Herweg and K. Rehkämper (eds.): *Raumkonzepte in Verstehensprozessen* . Tübingen: Niemeyer.

Kirsch, B., Schnepf, U., and Wachsmuth, I. (1993). RoboVis – a Scenario for Using Virtual Reality Techniques in Learning Robot Development, *this volume.*

Kozierok, R. and Maes, P. (1993). A learning interface agent for scheduling meetings. Proc. ACM-SIGCHI International Workshop on Intelligent User Interfaces, Florida, January 1993.

Krüger, W. (1993). Virtual Reality – Anwendungen in Wissenschaft, Technik und Medizin. *it – Sonderheft "Multi Media"*, April 1993.

Lang, E. (1989). Objektschemata und räumliche Konfiguration. In Ch. Habel, M. Herweg and K. Rehkämper (Hrsg.): *Raumkonzepte in Verstehensprozessen* (pp. 150-173). Tübingen: Niemeyer.

Laurel, B. (1990). Interface agents: Metaphors with character. In B. Laurel (ed.) (1990): *The art of human-computer interface design.* Addison-Wesley.

Levelt, W. (1986). Zur sprachlichen Abbildung des Raumes: Deiktische und intrinsische Perspektive. In H.-G. Bosshardt (Hrsg.): *Perspektiven auf Sprache* (S. 187-211). Berlin: DeGruyter.

Marcus, A. (1993). Future directions in advanced user interface design. In Thalmann, N.M. and Thalmann, D. (eds.): *Communicating with virtual worlds* (pp. 2-13). New York: Springer-Verlag.

Meinkoehn, J. and Knoll, A. (1993). Recent results in organization and performance evaluation for large sensor networks. *Int. Conf. on Intelligent Autonomous Systems (IAS-3)*, Pittsburgh.

Minsky, M.L. (1986). *The society of mind.* New York: Simon and Schuster, Inc.

Müller, J. and Siekmann, J. (1991). Structured social agents. In W. Brauer and D. Hernández (eds.): *Verteilte Künstliche Intelligenz und kooperatives Arbeiten.* 4. Internationaler GI-Kongreß Wissensbasierte Systeme. Heidelberg: Springer.

Newell, A. (1981). The knowledge level (Presidential address, AAAI-80). *AI magazine 2*(2), 1-20.Parnas 1972

Raab, F.H. et al. (1979). Magnetic Position and Orientation Tracking System (Polhemus). *IEEE Trans. AES 15*(5), 709-718.

Authors' Addresses

Ipke Wachsmuth, Faculty of Technology, University of Bielefeld, P.O.Box 100131, D-33501 Bielefeld, ipke@techfak.uni-bielefeld.de

Yong Cao, Faculty of Technology, University of Bielefeld, P.O.Box 100131, D-33501 Bielefeld, yong@techfak.uni-bielefeld.de

Limited Distribution Notice

KI-NRW

KI-NRW is a long-term cooperative research initiative on "Applications of Artificial Intelligence in North-Rhine-Westphalia". It is organized informally as an association of research groups from different universities and research institutes in the Federal State of North-Rhine-Westphalia, Germany.