

A Virtual Interface Agent and its Agency

Ipke Wachsmuth*, Britta Lenzmann*, Tanja Jörding**,
Bernhard Jung*, Marc Latoschik*, Martin Fröhlich*

* AI & Computer Graphics Lab
University of Bielefeld, Germany
{ipke,britta,jung,marc,martinfr}@techfak.uni-bielefeld.de

** Multimedia Technology Group
Dresden University of Technology, Germany
tj4@irz.inf.tu-dresden.de

Abstract

In the VIENA Project (“Virtual Environments and Agents”) we develop easy-to-use virtual environments for interactive design and exploration. We have modeled and implemented a synthetic human-like agent, Hamilton, that inhabits a simulated office environment and acts as an embodied virtual interface agent (VIA). To explore or change the simulated environment, people can instruct Hamilton by way of verbal input and simple hand gestures. Hamilton has a variety of functionalities which are put in effect by its agency, a multi-agent system. In mediating an instruction, invisible agents track exact object locations and colorings, and they negotiate alternative ways of acting. Hamilton’s agency is also able to adapt to individual users’ preferences during run time. As the VIA is present in the synthetic scene, users can take advantage of its anthropomorphic features, and they can choose to communicate with the agent from an external or an immersed view.

Introduction

In the VIENA Project (“Virtual Environments and Agents”) we develop easy-to-use virtual environments for interactive design and exploration (Wachsmuth & Cao 1995). To this end, we use specialized systems which act as “interface agents” between the user and the application system (Laurel 1990). Interface agents, for instance, were used to accumulate knowledge about tasks, habits, and preferences of their users to act on their behalf (Maes & Kozierok 1993). In our work we pursue a view of interface agents which use knowledge of the application itself and add comfort in human-computer interaction by relieving the user from technical detail (Wachsmuth, Lenzmann, & Jung 1995). In addition, we want interface agents to take better advantage of the ways humans communicate. Particularly the idea of “situated agents” is of interest to us, that is, to exploit the actual situation as a source of information in perceiving and manipulating the environ-

ment and communicating with cooperating partners¹. We are also led by ideas in (Marcus 1993) envisioning anthropomorphic agents to be involved in future user interfaces, and by work of (Badler *et al.* 1991; Webber *et al.* 1995) describing the use of natural language instructions for character animation of articulated figures. Finally, we found multiagent architectures of great advantage to our goal of building intelligent interfaces for complex applications.

In this paper we describe a “virtual interface agent” (VIA) which assists people in exploring, and changing the design of, virtual environments. As an articulated anthropomorphic agent which is visible in the scene (Figure 1), the VIA can encourage the use of natural



Figure 1: The virtual interface agent Hamilton

language, and it can be conceived either as a second “person” or as a personification of the user. With the help of the VIA we can place the user’s eye in the virtual environment. The VIA communicates and coop-

¹Situated communication is a focus theme in a long-term Collaborative Research Centre at the University of Bielefeld (SFB 360, “Situated Artificial Communicators”).

erates with a human user in an overlapping perceptual situation; it inspects the internal description of a scene which the user can experience by eye, so both of them – interface agent and user – communicate about scene details from their “point of view.” But whereas the user is likely to have a qualitative grasp of the scene, the VIA can track exact object locations, colorings, etc., and thus can mediate qualitative instructions of the user to produce quantitative changes in the scene description. The necessary functionality is realized by a multi-agent system described further below in the paper.

Hamilton, a virtual interface agent

We have modeled and implemented a synthetic human-like agent, Hamilton, that inhabits a simulated office environment and acts as an embodied virtual interface agent (VIA). When directed to do so, Hamilton can change its position and appearance in the following ways (Jörding & Wachsmuth 1996).

- **Moving and turning**

The agent can move in the horizontal plane and turn around its vertical axis. Gravity and collision resistance of solid bodies are taken into account.

- **Looking**

To look around, the head of the agent can turn left, right, up, and down within certain human-like boundaries.

- **Pointing gesture**

Hamilton can issue a pointing gesture with one arm extended and the index finger stretched. Before pointing to an object, first the agent’s body turns to the corresponding position.

- **Waving gesture**

As an answer to the input “hello”, the agent turns to the virtual camera, i.e. looks at the user, and a waving arm is seen for some seconds (Figure 2). These actions can also be evoked in response to a waving of the user (we use a simple data glove for this).

- **Changing the size**

Through the possibility of changing the body size of Hamilton, explorations become more flexible. Thus, we have provided instructions that cause a shrinking or growing of the agent.

- **Changing the perspective**

The user can also switch from an external view to an immersed view to look “through the eyes” of Hamilton.



Figure 2: “Hello”

Instructing the VIA

To explore or change the simulated environment, people can instruct Hamilton by way of verbal input and simple hand gestures. Instructions can be issued by typing or by voice input (speaker-dependent isolated words so far). They can make use of pointing gestures of Hamilton or of the user (by way of data glove; cf. Figure 3). Sample instructions for Hamilton are shown below.

1. *Hamilton come here*
2. *put a palmtree on the desk*
3. *make the left table green*
4. *much darker*
5. *put a computer on the green table*
6. *hello*
7. *turn the chair left*
8. *a bit more*
9. *put this <user pointing gesture> computer on the floor*
10. *put it back*
11. *Hamilton point to the green table*
12. *change the view*
13. *put the chair between you and there*
14. *Hamilton be smaller*
- ...

In order to process such instructions, the VIA needs knowledge about its environment, i.e., the current scene description but also information about object types and properties, reference schemes for spatial transformations, etc. This kind of knowledge is made



Figure 3: User pointing gesture

available by an underlying agency, the VIENA system, described in the next section².

Hamilton's agency

Hamilton has a variety of functionalities which are put into effect by its agency (see Figure 4), a multi-agent system. Instructions are issued via a multimodal input agency that combines input from different modalities. It consists of input listeners, a parser, and a coordinator. Three listener agents, i.e., the type listener, speech listener, and gesture listener, track and analyze sensor data from the keyboard, the microphone, and the data glove, respectively. With the help of the parser, the coordinator analyzes and integrates the inputs received from the listeners and generates an internal task description that is posted to appropriate agents of Hamilton's agency.

In mediating an instruction, invisible agents in the VIENA system track exact object locations and colorings, and they negotiate alternative ways of acting. For example, a space agent computes spatial transformations in the virtual environment such as translating, rotating, and scaling of scene objects. By inspecting or modifying (r,g,b)-vectors, a color agent helps to identify an object by means of a color description ("the red chair") or to change the appearance of objects (e.g., blue, lighter). A camera agent calculates transformations of the virtual camera to enable the navigation through the scene. To resolve ambiguous references in the qualitative instruction, a reference agent identifies an individual scene object. A Hamilton control agent

realizes the manipulation of the articulated figure. A bookkeeping agent is authorized to access and modify the augmented graphics data base to supply current situation information to agents on request. Agents' functionalities account for implicit assumptions of the human when manipulating the virtual environment; physical laws were reconstructed to avoid the collisions of massive objects in the virtual world; cognitive factors of space are recognized when converting instructions to scene alterations.

A stand-alone MACE-type framework (Gasser, Braganza, & Herman 1987) was developed which standardizes the communicative and cooperative behaviors of all agents. Agents cooperate to offer a goal scene corresponding to a user's instruction. The offer can be changed in further interaction, that is, the user can negotiate the semantics of instructions.

Adaptation to users' preferences

Hamilton's agency is also able to adapt to individual users' preferences during run time. Since the practical experience with the VIENA system has shown that significant variations of users' preferences exist with respect to possible solutions, we have built in an adaptation method (Lenzmann & Wachsmuth 1996).

For example, to adapt to users' preferences for different spatial reference frames, we have implemented different instances of the space agent. These instances are alike in the way they compute spatial transformations, and different with respect to the reference scheme they assume. We have implemented one space agent based on the user's egocentric reference frame (deictic reference) and one space agent with an externally anchored reference frame (intrinsic reference). Similarly, we have implemented two instances of agents controlling Hamilton which compute Hamilton's spatial movements on the basis of alternative reference frames. We have also realized two color agents that offer more drastic or smoother color transformations to adapt to users' preferences for different color sensation.

In case a visualized solution does not meet the user's expectation, the user can correct the system by stating 'wrong'. The negative user feedback leads the agents to reorganize themselves in a way that one of the other agent instances generates an offer which modifies the previous solution.

To this end, agents use an informed negotiation process. Knowledge about the user and the preceding session is captured in internal credit vectors. By learning from indirect user feedback, agents compete with each other to meet the users' preferences. Thus, the system's knowledge of the users is expressed in the activation of certain agents in the entire interface agency.

²A running demo of an earlier version of the VIENA-System is part of our contribution to the IJCAI-95 Videotape program.

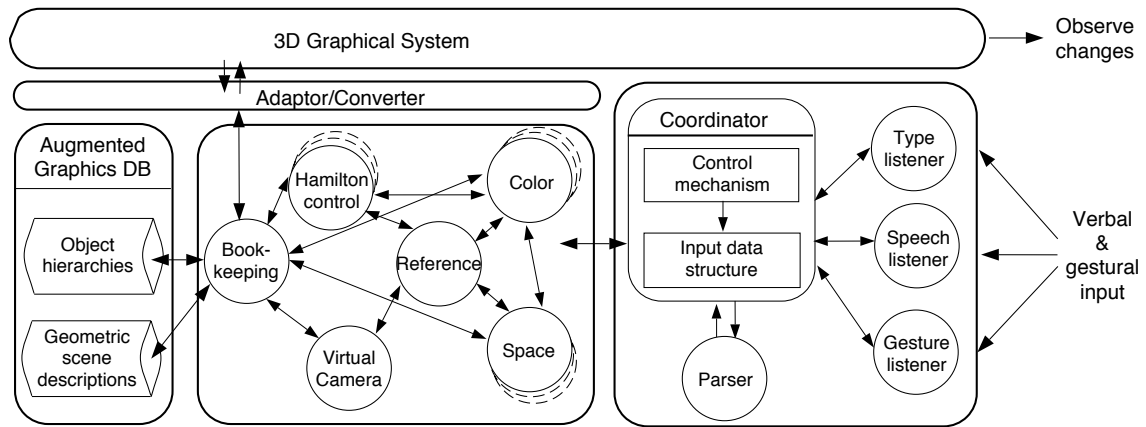


Figure 4: Hamilton's agency mediates qualitative verbal and gestural instructions by translating them to quantitative commands that are used to update the visualization scene model.

In this way, user adaptation is achieved without accumulating explicit user models.

Anthropomorphic features in communication

As the VIA is present in the synthetic scene, users can take advantage of its anthropomorphic features, and they can choose to communicate with the agent from an external or an immersed view (Jörding & Wachsmuth 1996). Thus, the user has two possibilities for exploring the virtual environment:

- in the **external view** the agent is visible in the scene and can serve as an anchored reference frame in communication. The user can direct the agent to move in the scene and assess ergonomic features of the environment (e.g. size of a table) in comparison to anthropometric features of the human-like synthetic agent; cf. Figure 5.
- in the **immersed view** the user adopts the same perspective and field of vision as the anthropomorphic agent to feel more involved in the scene. Figure 6 shows the view “through the eyes” of Hamilton during a pointing gesture.

Because of the different perspectives the user can assume, it has to be clarified from which perspective the user has issued an instruction. In the external view, the user could take on the view of the virtual camera which determines the current field of vision. Alternatively, the user could take on the view of the anthropomorphic agent and identify the position of the agent with his/her own. For instance, in response to the instruction *move the chair here*, the chair would be moved either near the anthropomorphic agent or



Figure 5: Pointing (external view)

near the virtual camera, that is, toward the front of the screen. If the user changes to the immersed view, these two perspectives coincide. Similarly, *left* and *right* have different meanings, depending on the current perspective.

A bit more complicated is an instruction like *move the chair there*. The verbal expression *there* is most often combined with a gesture. In the VIENA system the anthropomorphic agent can carry out a pointing gesture with its right arm, or the user can point with the data glove. In the presence of a pointing gesture, *there* can indicate a region in the direction of the pointing. Another clue for locating the position *there* can be the line of sight of the user. When having the immersed view, the user can move in the virtual room looking “through Hamilton’s eyes”, and can shrink or

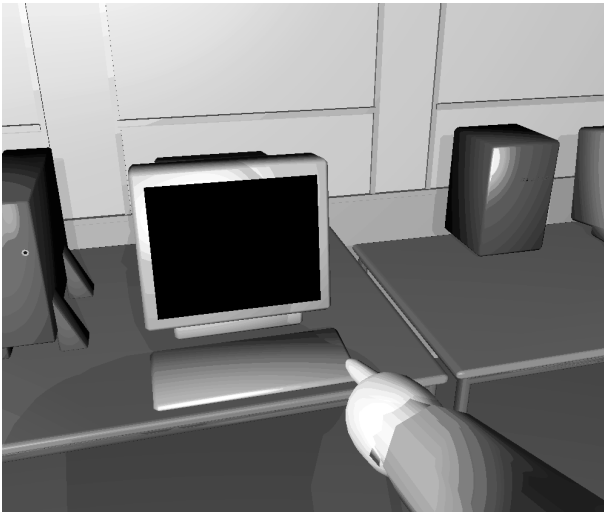


Figure 6: Pointing (immersed view)

grow to explore the environment from the perspective of a child or a tall person.

Discussion and Future Work

In this paper, we presented an approach that integrates ideas from synthetic articulated agents and from software interface agents, leading into what we call a virtual interface agent (VIA). We described an application where a VIA is used as an intelligent mediator in communicating with a virtual environment. In our example domain of interior design, the VIA assists in exploring and changing the design of the environment. By being situated in its environment, the VIA has the ability to exploit the actual situation as a source of information in communicating and collaborating with a human user in an overlapping perceptual situation. The VIA also serves as an anchored reference system which allows the use of spatial language in communication, based on a variety of reference schemes. By this, natural interaction with the environment is greatly enhanced.

Using a multi-agent system as an invisible part of the VIA, besides being justified from using distributed resources, has a number of further advantages. For instance, it provides a good means to process multimodal input: Different modes of interaction can be integrated by the cooperation of agents. A further aspect in our work is adaptivity: Agents making offers qualified by their success in the preceding session provide the basis to adapt the interface system to individual users' preferences without accumulating user models. We have started to incorporate anytime facilities, i.e., crude solutions can be offered by simple agents early, while

more fine-tuned agents still work on a better solution as long as allocated compute time has not elapsed.

Besides of relieving technical detail from the user by making an intelligent agency do part of the work, incorporating language and gestures as an interaction means seems an important aspect.

Much work in human-computer interaction for virtual environments has concentrated on gesturing and pointing; e.g., (Böhm, Hübner, & Väänänen 1992; Wexelblat 1995). But when interacting with synthetic human-like agents, it seems natural to also include means of verbal interaction. In the VIENA Project we have started out to use natural language instructions. On the one hand, it is often easier to instruct the system to carry out changes when gestural manipulation is impossible or unnatural ("make the table green"). On the other hand, gestural interaction is especially valuable to convey spatial information in an intuitive way. We will now place more emphasis on co-verbal gestures. We have started to work on automatic gesture transcription to encode user's movements into symbols for further interpretation in the context of language instructions. We also want to make use of iconic and pantomimic gestures to take advantage of a wider range of gestural expression in communicating with the interface agency.

In our next work, on multimedia interfaces, we will change to a setting where a virtual environment is presented on a large-screen display. In this setting, the synthetic agent is close to human-size, and much finer-grained details in the environment will be recognized in interaction. We think that synthetic agents can provide means for easy-to-use virtual environments and for building new ways of interface technology.

Acknowledgments

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 authors are indebted to Peter Dawabi, Carla Intrup, Marko Merkle, Ralf Nolte and Timo Sowa for assistance in the research.

References

- Badler, N. I.; Webber, B. L.; Kalita, J.; and Esakov, J. 1991. *Animation from Instructions*. San Mateo, (Cal.): Morgan Kaufmann. 51–93.
- Böhm, K.; Hübner, W.; and Väänänen, K. 1992. Given: Gesture driven interactions in virtual environments: A toolkit approach to 3D interactions. In *Proc. Conf. Interface to Real and Virtual Worlds*.
- Gasser, L.; Braganza, C.; and Herman, N. 1987. Implementing distributed AI systems using MACE.

In Bond, A., and Gasser, L., eds., *Readings in Distributed Artificial Intelligence*. Morgan Kaufmann. 445–450.

Jörding, T., and Wachsmuth, I. 1996. An anthropomorphic agent for the use of spatial language. Accepted for: ECAI-96 Workshop “Representation and Processing of Spatial Expressions”.

Laurel, B. 1990. Interface agents: Metaphors with character. In Laurel, B., ed., *The art of human-computer interface design*. Addison-Wesley. 355–365.

Lenzmann, B., and Wachsmuth, I. 1996. A user-adaptive interface agency for interaction with a virtual environment. In Weiss, G., and Sen, S., eds., *Adaption and Learning in Multi-Agent Systems*, LNAI 1042. Berlin: Springer. 140–151.

Maes, P., and Kozierok, R. 1993. Learning interface agents. In *Proceedings of the Eleventh National Conference on Artificial Intelligence (AAAI-93)*, 459–465. AAAI Press/The MIT Press.

Marcus, A. 1993. Future directions in advanced user interface design. In Thalmann, N., and Thalmann, D., eds., *Communicating with Virtual Worlds*. New York: Springer-Verlag. 2–13.

Wachsmuth, I., and Cao, Y. 1995. Interactive graphics design with situated agents. In Straßer, W., and Wahl, F., eds., *Graphics and Robotics*. Springer. 73–85.

Wachsmuth, I.; Lenzmann, B.; and Jung, B. 1995. Communicating with virtual environments - a survey of recent work at the University of Bielefeld. In Feller, D. W., ed., *Modeling - Virtual Worlds - Distributed Graphics MVD'95*, 93–97. Sankt Augustin: infix.

Webber, B.; Badler, N.; Di Eugenio, B.; Geib, C.; Levison, L.; and Moore, M. 1995. Instructions, intentions and expectations. *Artificial Intelligence* 73(1-2):253–269.

Wexelblat, A. D. 1995. An approach to natural gesture in virtual environments. *ACM Transactions on Computer-Human Interaction (TOCHI)* 2(3):179–200.