Towards Communicating Agents and Avatars in Virtual Worlds

Anton Nijholt and Hendri Hondorp [†]

University of Twente,
PO Box 217, 7500 AE Enschede,
the Netherlands
email: {anijholt|hendri}cs.utwente.nl

Abstract

We report about ongoing research in a virtual reality environment where visitors can interact with agents that help them to obtain information, to perform certain transactions and to collaborate with them in order to get some tasks done. In addition, in a multi-user version of the system visitors can chat with each other. Our environment is a laboratory for research and for experiments with users interacting with agents in multimodal ways, referring to visualized information and making use of knowledge possessed by domain agents, but also by agents that represent other visitors of this environment. We discuss standards that are under development for designing such environments. Our environment models a local theatre in our hometown. We discuss our attempts to let this environment evolve into a theatre community where we do not only have goal-directed visitors buying tickets, but also visitors that that are not yet sure whether they want to buy or just want information or visitors who just want to look around, talk with others, etc. It is shown that we need a multi-user and multi-agent environment to realize our goals and that we need to have a unifying framework in order to be able to introduce and maintain different agents and user avatars with different abilities, including intellectual, interaction and animation abilities.

1. Introduction

We discuss a virtual reality theatre environment in which we have embedded agents that can help the user through natural language dialogue. The environment has been built using VRML (Virtual Reality Modeling Language) and is accessible on WWW. Originally the environment was built around an existing natural language dialogue system allowing dialogues about performances and reservations for theatre performances [Lie et al.[6]]. In the environment the system has been assigned to a visualized embodied agent to which users can ask questions. Once we had this agent and extended the environment there grew the need to add other agents that were able to help the visitor, that were able to communicate with each other and that were able to show some autonomous behavior. We discuss how our ideas about this environment changed in time, in particular by paying more attention to potential users. Rather than a goal-directed information and transaction system comparable with a voice-only telephone information system, the environment is now evolving into a virtual community where differences between visitors and artificial agents become blurred and where research topics range from assigning personalities and emotions to artificial agents, usability studies involving a navigational assistant to formal specification of (interactions in) virtual environments and reinforcement learning for agents in this virtual, multimodal environment in order to increase an agent's autonomy.

2. Building the Virtual Environment

In Figure 1 we have an aerial photograph of the centre of Enschede. It includes the market square, the old church (notice its black shadow in the middle of the photograph) and some theatre buildings. The main theatre building is on the right. It is called the 'MuziekCentrum'. It includes some performance halls, rooms for artists, recreational locations

[†] Research reported in this paper has been made possible by the "VR Valley Twente" foundation and by the U-Wish project of the Dutch Telematics Institute.

Figure 1: *Aerial View of the City of Enschede*

(for audience and performers), wardrobes, etc. It also includes a conservatory. There are other theatre buildings in this town. Information about performances can be obtained by the usual brochures, advertisements and announcements in newspapers, by phone (including a *press 3 if you want more information about* ... dial system) and by simply taking the bike, go to one of the theatre buildings and ask a receptionist about performances and then make a reservation for a particular performance.

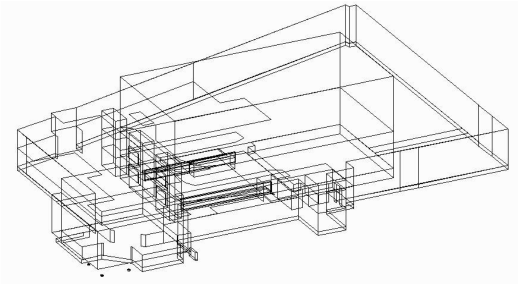

Figure 2: *Wire Model of the Theatre*

At this moment some of the theatre buildings, their environment and the streets leading from one location to the other have been modeled in VRML. In Figure 2 we have a wire model of the main theatre. The theatre was built according to design drawings of the architects of the building. Rooms, stairs, stages, etc., according to the actual building were added and textures using photographs and video of the real building were glued to objects, walls, ceilings, etc. in order to make the virtual environment realistic. We did not pursue a hundred percent realism. Obviously, we have to deal with time limits, availability of students and availability of programmers who can work on our environment. However, there is not always a need to strive for complete realism. In a virtual environment we can give indications to visitors how to achieve certain goals without being bothered by, e.g., physical or social constraints.

In Figure 3 we have a screenshot taken from the entrance of the virtual MuziekCentrum, the real local theatre building that we converted into a virtual 'local' theatre. In this screenshot the doors are open, we see part of the environment and when we look inside we see an information desk and the Karin agent waiting to tell us about performances, artists and available tickets.

Visitors can explore this virtual environment, walk from one location to another, ask questions to visible agents, click on objects, etc. Karin, the receptionist of the theatre, has a 3-D face that allows simple facial expressions and lip movements that synchronize with a text-to-speech system that mouths the system's utterances to the user. Because of web limitations, there is no sophisticated synchronization between the (contents of the) utterances produced by the dialogue manager and corresponding lip movements and facial expressions of the Karin agent. Design considerations that allow an embodied agent like Karin to display combinations of verbal and non-verbal behavior can be found in [Nijholt/Hulstijn[7]].

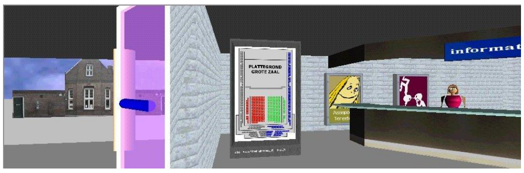

Figure 3: *View from the Entrance of the Inside*

Multimodality is another issue. Karin decides to present a table on the screen when there are too many performances she has to read. Clearly, when there are too many performances that satisfy the user's requirements we can not expect that after Karin has read the information about the third or fourth performance the user still knows the details of the first performance. Therefore we decided to embed Karin and her information desk in a windows environment where we can also have the possibility to show information in tables with clickable items and pop-up menu's of frequently asked questions. In the dialogue system it has been made possible to make references to the items in the table. That is, instead of clicking one of the frequently asked questions, it is as well possible to ask a question like: *"Please give me more information about the third performance"*, making a reference to the third item in the table of available performances.

Other agents in this environment have been introduced. For example, there is a navigation agent which knows about the geography of the building and can be addressed using speech and keyboard input of natural language. No real dialogues are involved. The visitor can ask about existing locations in the theatre. When recognized, a route is computed and the visitor's viewpoint is guided along this route to the destination. The navigation agents has not been visualized as an avatar. Its viewpoint in the theatre is the current viewpoint from the position (coordinates) of the visitor in the world. A Java based agent framework has been introduced to provide the protocol for communication between agents. It allows the introduction of other agents. For example, why not allow the visitor to talk to the theatre seat map or to a poster displaying an interesting performance? Unlike its predecessor, the version of the virtual theatre with a speech recognizing navigation agent has not been made accessible to the general audience by putting it on the Web. Although speech recognition is done at the server (avoiding problems of download time, ownership, etc.) there are nevertheless too many problems with recognition quality and synchronization with the events in the system. However, further work on the navigation agent is in progress. Part of this work is on user preferences for navigation in virtual worlds, part is on modeling navigation knowledge and navigation dialogues, part is on adding instruction models to agents and part is on visualization.

3. Visitors that Interact with Avatars

In our environment we can have different human-like agents. Some of them are represented as communicative humanoids, more or less naturally visualized avatars standing or moving around in the virtual world and allowing interaction with visitors of the environment. In a browser which allows the visualization of multiple users, other visitors become visible as avatars. We want any visitor to be able to communicate with agents and other visitors, whether visualized or not, in his or her view. That means we can have conversations between agents, between visitors, and between visitors and agents. This is a rather ambitious goal which can not be realized yet.

An other problem we should mention is that communication situated in a visible or otherwise observable (virtual) shared environment allows the communicating partners to support there communicative acts by other means of directing (like gazing or pointing) than linguistic reference. Introducing this multi-modal support for language communication in some cases helps the agents to understand each other but it introduces some new and challenging problems as well. One of them is the problem of coreferencing to shared visible objects. The phrase *'that door'* should be attached to some visible object in the environment and assumes that the agents share the visibility of this object. The 'geometrical' virtual environment (described in VRML code or in whatever virtual modeling language) must be described on an abstract conceptual and linguistic level as well. The agent should somehow be able to know what object the user points at even in case it is not in direct view of the agent and it should therefore be able to match this way of referring with the linguistic reference ("that door").

In the previous sections we talked about agents acting in our own virtual theatre. Karin was introduced as a 'visualization' of our existing dialogue system. She has extensive knowledge of performances that play in the theatre. She can move her lips and have some simple head movements in function of the dialogue. Once we had Karin it became clear that we needed an agent framework and in it we introduced a navigation agent with some geographical knowledge and speech recognition capabilities. In fact, we have a multitude of potential and useful agents in our environment, where some just perform some animation, others can walk around (e.g., this would be useful for a navigation agent) and others have some built-in intelligence that allows them to execute certain actions based on interactions with visitors.

4. Multi-agents and Multi-users

We embedded our environment in a multi-user shell (DeepMatrix [Reitmayer et al.[8]]) which means that visitors become visible as avatars (VRML objects) to which we can assign animations, but also intelligence and interaction abilities which can reflect those of the visitor, but not necessarily, since we can modify them to suit our purposes (or application). As an example, we can have user profiles (obtained by learning, by assuming or by asking) assigned by the system to the visitor's avatar acting in the virtual environment. In this way, in an E-commerce environment we can get track of and anticipate different consuming buying behavior [Guttman et al.[5]] by reading the visitor's user profile.

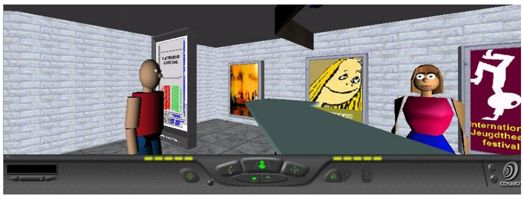

Figure 4: *Jacob Talking to Karin*

In Figure 4 we see a visitor's avatar approaching Karin. The visitor has chosen one of our avatars (Jacob) that knows to walk around in the virtual theatre. Its animations allow it to walk by following the coordinates of the moving viewpoint position of its owner. Jacob has been introduced in one of our other projects (see [Evers and Nijholt[3]]) in which it has been assigned an instruction and task model to teach a particular task. That is, Jacob knows about the Towers of Hanoi, how to teach this problem to students and how to interact with students about this problem. Presently, Jacob is only an example how we can reuse VRML objects for physical appearances and animations of agents when we comply to standards. There is no way to translate Jacob's intelligence to a different VRML environment. Standards need to be developed that deal with 'intelligence' issues of agents in addition to issues that deal with size, appearance and animations in VRML worlds.

In our environment there exist more possibilities for a user's avatar to meet artificial agents. For example, we have a piano player on stage with some simple predefined animations accompanying the music. At the Università degli Studi di Milano research has been done on baroque dance animation with virtual dancers [Bertolo et al.[1]]. Using a baroque dance editor dances performed by virtual dancers can be choreographed and generated. Since the generated dances and animations are described in VRML it has become possible to have some guest performances of the Scala of Milan dancers in our theatre. In Figure 5 we also see that the visitor represented by the Jacob avatar has been so impertinent to climb the stage in order to get a closer look at the performing dancer. It will be clear that in order to maintain a virtual environment where we have a multitude of domain and user-defined agents we need some uniformity from which we can diverge in several directions and combinations of directions: agent intelligence, agent interaction capabilities, agent visualization and agent animation.

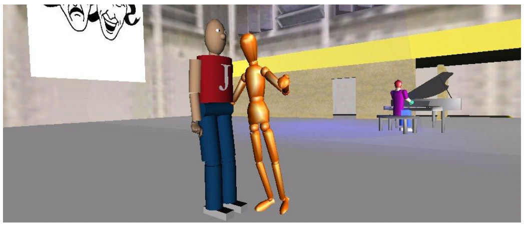

Figure 5: *Jacob in Conflict with the Baroque Dancer*

We can look at some VRML related standards that have been proposed or are under development. For our aims, we are interested in:

- Humanoid Animation (H-Anim) standard [VRML Humanoid[9]]. This standard defines a structure and interface for agents in VRML. An agent that conforms to the standard can be plugged into a VRML world and controlled through its interface. Animations can be added to the H-Anim agents.
- MPEG-4SNHC standard, a set of definitions including a set of facial animation parameters and a set of feature points for facial components [VRML-MPEG4[10]].
- Living Worlds Standard. The aim is to define a conceptual framework and specify interfaces to support the creation of multi-user and multi-developer applications in VRML [VRML Living Worlds[11]]. Standards should allow applications which support the virtual presence of many people in a single scene at the same time: people who can interact with objects in the scene and with each other. Moreover, they allow that applications can be assembled from libraries of components developed independently by multiple suppliers.

Jacob has been built following the H-anim standard. As mentioned, presently we use the DeepMatrix multi-user environment system. It is compliant with the Living Worlds specification. This specification deals with data distribution and scene synchronization. Below this are standards dealing with network and application protocols. Beyond the Living Worlds specification are the issues which have to be dealt with in order to introduce standardized interacting agent frameworks in virtual environments.

5. Conclusions and Future Research

In conclusion, we think that for our environment the following three lines of research have to be taken simultaneously:

- Redesigning and extending our agent framework such that individual agents can represent (human) visitors (e.g., movements, posture, nonverbal behavior) and can stand for artificial, embodied domain agents that help visitors in the virtual environment (using multimodal interaction, including speech and language).
- Designing H-Anim agents that are controlled according to the protocol of the agent framework, that can walk around in the virtual environment (either acting as a domain agent, hence displaying intelligent and autonomous behavior, or representing a visitor and its moving around in the environment).
- Relating the agent framework to the theory of multi-agent systems and issues of autonomy, reactivity, pro-activity, social ability and learning. General frameworks for intelligent agents have been developed, among them the theory of belief-desire-intention agents.

In a multi-user environment it has to be decided which parts of the environment are shared and which parts are 'private', that is, parts in which events are only noticeable for one user, not leading to updates of the same part when being visited by other users. Environments in which we have interactions with the world and with agents in this world account for many interesting problems. For example, has every visitor of the environment its 'own' Karin? Or do they have to queue in order to get their turn to speak to her? In Figure 3 we see that Karin is looking to an other visitor (not visible) rather than to Jacob. Non-verbal behavior, including gaze modelling, becomes important when we have different visitors engaged in a conversation with an agent. In [Vertegaal et al.[12]] we report about experiments and a prototype version of an environment in which such conversations take place.

As an other example of the problems involved when we share an environment with others, consider our theatre hall when we decide to have some performance on the stage. As mentioned in [Reitmayer et al.[8]], having a large crowd as audience introduces all the real world logistics of event presentation, including seat assignment and sight lines. This is in fact the situation described in Neal Stephenson's *Snow Crash* (1993) where hundred thousands of hackers represented as avatars fill an amphitheater to watch a performance. An alternative would be to enter a performance hall without sharing this environment with others, hence, being able to take the best seat (as many others will), to move on-stage as done by Jacob in Figure 5, etc.

As a final conclusion we want to emphasize once more the necessity to have standards. However, this is becoming a rather accepted issue. Examples of agent communication standards that have been introduced in order to obtain interoperability are FIPA (Foundation of Intelligent Physical Agents: http://www.fipa.org) and KQML (Knowledge Query and Manipulation Language). The FIPA specification allows the construction and management of an agent system composed of different agents, possibly built by different developers. It specifies how agents can interact with humans, other agents, non-agent software and the physical world. The FIPA Agent Communication Language (ACL) is based on speech act theory. This seems to be a useful starting point for relating these formal communication languages with the use of natural speech and language in dialogue systems between visitors and artificial agents [Dahlbäck et al.[2]] or in mediated communication between visitors of virtual environments. Attempts to define standards for dialogue systems can be found in [Gibbon et al.[4]]. Recently DARPA has adopted for its Communicator Program on spoken dialogue systems a dialogue system architecture, the MIT Galaxy System, as the reference architecture for dialogue research groups that participate in DARPA projects (http://fofoca.mitre.org). It is assumed that in this way more commercial standards will emerge in the speech and language areas.

It is clear that in order to design and implement virtual environments that are inhabited by multiple agents and users and that are developed and extended by different developers it is necessary to comply to standards like the ones mentioned here.

References

1. M. Bertolo, P. Maninetti and D. Marini. *Baroque dance animation with virtual dancers. Eurographics '99, Short Papers and Demos*, Milan, 1999, 117-120. 4
2. N. Dahlbäck, N. Reithinger and M.A. Walker. *Standards for dialogue coding in natural language processing.* Report on a Dagstuhl-Seminar, February 1997. 5
3. M. Evers and A. Nijholt. *Jacob - An animated instruction agent in virtual reality.* Submitted for publication. 3
4. D. Gibbon, R. Moore and R. Winski. *Handbook of Standards and Resources for Spoken Language Systems*. Mouton de Gruyter, Berlin, 1997. 5
5. R.H. Guttman, A.G. Moukas and P. Maes. *Agent-mediated electronic commerce: a survey.* Knowledge Engineering Review, June 1998. 3
6. D. Lie, J. Hulstijn, R. op den Akker and A. Nijholt. *A Transformational Approach to NL Understanding in Dialogue Systems.* Proc. NLP and Industrial Applications, Moncton, 1998, 163-168. 1
7. A. Nijholt and J. Hulstijn. *Multimodal Interactions with Agents in Virtual Worlds.* In: Future Directions for Intelligent Information Systems and Information Science, N. Kasabov (ed.), Physica-Verlag: Studies in Fuzziness and Soft Computing, 2000. 2
8. G. Reitmayr, S. Carroll, A. Reitemeyer and M.G. Wagner. *Deep Matrix: An open technology based virtual environment system.* The Visual Computer Journal 15:395-412, 1999. 3, 4
9. VRML Humanoid Animation Working Group, http://ece.uwaterloo.ca/ h-anim, 1998. 4
10. VRML-MPEG4 Working Group, http://www.vrml.org/WorkingGroups/vrml-mpeg4. 4
11. VRML Living Worlds Working Group: *Making VRML 97 Applications Interpersonal and Interoperable.* http://www.vrml.org/WorkingGroups/living-worlds, 1998. 4
12. R. Vertegaal, R. Slagter, G. van der Veer and A. Nijholt. *Why conversational agents should catch the eye.* Proceedings CHI 2000: Extended Abstracts, 2000, 257-258. 4