

3. From Virtual Environment to Virtual Community

A. Nijholt

University of Twente, Department of Computer Science, PO Box 217, 7500 AE Enschede, Netherlands

3.1 Introduction

We discuss a virtual reality theater environment and its transition to a virtual community by adding domain agents and by allowing multiple users to visit this environment. The environment has been built using VRML (Virtual Reality Modeling Language). We discuss how our ideas about this environment changed in time by adding more facilities to it and by paying more attention to potential users. Rather than a goal-directed information and transaction system, the environment is evolving into a virtual community where differences between visitors and artificial agents can become blurred. Before going into a description of our own environment and its development we survey the research areas that now allow the building of 3D embodied and animated agents that show intelligence and personality and that can inhabit our environment.

3.2 Towards Multi-user Virtual Worlds

The first networked virtual worlds were text-based. They became known as MUDs (Multi-User Domains) and they allowed communication between users and access to a shared database with text descriptions of users and objects. In these environments the personality of a user shows in the contents and the style of the text utterances the user produces, his turn taking behavior and more generally the moods (as they show) and attitudes towards the community that can develop in such environments. Graphical multi-user environments were introduced in the 1980s. In a typical setting we have a background image showing the entrances to several locations or rooms in the environment or we are in one of these 2D locations and we can choose one of the other visitors (or all of them) to talk to. Typically, visitors can present themselves by choosing an avatar (a 2D object) and its predefined animations. These animations are simple (a waving gesture, a jump of joy, . . .). Most interactions are text-based, by using chat windows and text balloons that appear above the head of avatars that take part in the discussion.

With the advent of VRML, virtual worlds could be designed for Worldwide Web. Rather than for chatting, the worlds were meant to be explored, to explain or to allow the simulation of a particular activity in which the visitor was involved. Virtual reality applications were already there and rather than consider distributed virtual reality as a technology to design communities it was explored for all kinds

of applications. Virtual worlds intended to meet other people entered the arena. In these worlds multiple visitors can share the scenes. In the more advanced worlds users can change parts of the world and can have sophisticated visual representations that can interact without being restricted to predefined gestures. An avatar can be made to resemble the human user by photographic means.

The worlds that we consider may have collision and gravity features that become visible in the movements of avatars. There can be real-time voice communication and in addition there can be lip-sync facial gestures. Despite adding such features, there remains an enormous gap when we compare the capabilities of the avatars and talking heads with those of the humans they represent. One way to close this gap is to give the human user the ability to control the avatar in a much more detailed way. One possibility is to have them explicitly controlled online by the user and captured from verbal and non-verbal input or from body movements. Also, in addition to the avatars that represent humans we can add domain avatars to the environment to increase the sense of reality. They should be animated, but preferably there should be possibilities to give them personality and capabilities to act on their own or on behalf of a user of the avatar or owner of the environment. That is, they need appropriate internal modeling to allow autonomous behavior.

3.2.1 Interacting Embodied Personalities

Agent technology is a research field that emerged in the 1990's and that can be considered as a field in which actors are developed, although not necessarily in the context of human-computer interaction or virtual communities. Without going into details and especially controversial details, we want to mention properties of software modules that are generally assumed to be present before being 'allowed' to talk about them as agents: autonomy, reactive and proactive behavior and the ability to interact with other agents (or humans). For an agent to act appropriately in a domain it has been useful to distinguish beliefs (what the agent regards to be true, this may change in time), desires (the goals the agent has committed himself to) and the intentions (short-term plans that it tries to execute).

Believability is a notion that has been emphasized by Joseph Bates, again in the early 1990's. An agent is called believable, if some version of a personality shows in the interaction with a human. Two main theories on personality which can be used to design believable agents are trait theory, where personality is a set of psychological traits that characterizes a person's behavior and social learning theory, where appraisal of the situation and the individual's history are taken into account. Main requirements for believability are (Loyall [8]): personality, emotion, self-motivation, change, social relationships and consistency of expression.

When we zoom in on the role of emotions, it should be mentioned that there are many subtleties involved when conveying them. Cartoon characters are allowed to exaggerate, giving more cues to the observer. Emotional cues shouldn't be in conflict with contextual cues. Emotional cues should be consistent during interaction; nevertheless they may change when interaction has taken place with the same user during a longer period, in time. Computational models from which emotional behavior can be generated exist, but are not based on well-developed theory. There-

fore, rather than having emergent emotional behavior based on an agent's cognitive appraisal model, we see applications in prototype (learning) environments with preprogrammed emotional display.

Now that we have discussed reasonable, social, intelligent, believable and, indeed, whatever kind of cognitive behavior, it is time to consider the role of embodiment. Embodiment allows more agent multimodality, therefore making interaction more natural and robust. Several authors have investigated nonverbal behavior among humans and the role and use of nonverbal behavior to support human-computer interaction. See e.g. (Cassell [1]) for a collection of chapters on properties and impact of embodied conversational agents (with an emphasis on coherent facial expressions, gestures, intonation, posture and gaze in communication) and for the role of embodiment (and small talk) on fostering self-disclosure and trust building. While the previous investigations we mentioned can be understood to emphasize the cognitive viewpoint of embodiment, we can as much emphasize the possibility of an embodied agent to walk around, to point at objects in a visualized domain, to manipulate objects or to change a visualized (virtual) environment. In these cases the embodiment can provide a point of the focus for interaction. From a technical point of view, extremely much has to be done on human-like (from a physical and cognitive point of view) agent behavior. From a domain point of view it has to be decided when and why such behavior is useful.

Our next step is from embodiment to virtual humans. A list of research topics involved includes natural looking movement and deformation of visible body surface, animation of skeleton, hands and face, hair, skin and clothes representation, natural looking walking and grasping animation and, very importantly in the view of the previous topics, behavioral animation which strives at giving character and personality to the animation. This list of viewpoints can be complemented with viewpoints from cognitive and perceptory sciences. Virtual humans have to act in virtual environments where a visual, an auditory and a haptic/kinaesthetic environment intersect.

3.2.2 Embodied Personalities in Virtual Worlds

Agents are finding their way in virtual environments. The first applications of embodied agents can be found in training, simulation, education and entertainment. These environments may include a single agent with which the user can interact, but the user itself, or part of the user, can be included in the environment. In team training we can have several agents in the environment or several users are represented in the environment. Research into crowd modeling also studies the behavior of groups of people in virtual environments.

However, apart from these applications we also see developments where 2D and 3D extensions of chat worlds and digital cities become inhabited by embodied agents, both as representations of visitors and as autonomous domain agents. In the near future we can expect that companies, families or groups of people that share interests have the opportunity to design and use such environments. Below we mention a few projects in which these future developments become visible.

Several impressive research systems employing animated pedagogical agents have been built and are in a process of further development. Embodied pedagogical agents can show how to manipulate objects, they can demonstrate tasks and they can employ gesture to focus attention. As such they can give more customized advice in an information-rich environment. Lester et al. [7] use the term deictic believability for agents that are situated in a world that they co-inhabit with students and in which they use their knowledge of the world, their relative location and their previous actions to create natural deictic gestures, motions, and utterances.

One example of an environment that employs embodied agents is the Soar Training Expert for Virtual Environments (STEVE, see Johnson et al. [5]). This is an immersive 3-D learning environment with a virtual agent called Steve. Steve demonstrates how to perform a physical, procedural task. It is a typical example of an environment where a student can get hands-on experience. Due to the student's head-mounted display, Steve's perception module knows about the student's position in the virtual world, about the student's line of sight and which objects are in the student's field of view. Steve has been designed to support team training.

A second example we want to mention is a BodyChat (Vilhjalmsson [14]), a research environment on conversational embodied agents. That is, there is not really a task to be performed or learned. People exchange information and chat. In this environment several users can have a conversation using the keyboard while their cartoon-like 3D animated avatars display corresponding salutations and turn taking behavior. They look away during planning an utterance, they back-channel feedback and facial expression and look to the next speaker when ending. Watanabe [15] reports about similar research. Another system by Vilhjalmsson, called Situated Chat is in development. This system also animates avatars in an online graphical chat environment. However, since it knows about the shared visual environment the generation of avatar movements can include referring gestures when making implicit or explicit references to the environment during the conversation.

As a third range of examples we look at systems that have become known as interactive theater, where players connected by a network can take part in a performance as actors. There is a host server for the producer and there are client computers for the performers. The latter are represented as avatars in the virtual environment and with motion capture systems (cameras or sensors) avatar movements reflect player actions. Gestures, touch and facial expressions of the players can be tracked and given to the animation algorithms. The virtual stage may have actors that are provided by the theater and that show autonomous behavior according to some action patterns. They have a role, but the way they perform this role is also determined in interactions with the human players and their alter ego avatars. See Takahashi et al. [12] and Tosa et al. [13] for examples of interactive theater.

3.3 Building a Theater Environment

The main theater building in our university town is called 'Het Muziek Centrum'. It includes the usual rooms: performance halls, dressing rooms for artists, recreational locations (for the audience and performers), wardrobes, etcetera. It also includes a music academy. There are also some other theater buildings in the town. At this moment some of the buildings, their surroundings and the streets leading from one location to the other are being modeled in VRML and Java 3D. The virtual theater was built according to the design drawings of the architects of the real building. Originally the environment was built around an already existing natural language dialogue system that provides information about theater performances and that allows reservations to be made. In the virtual environment the dialogue system has been assigned to a visualized embodied agent. Once we had this agent and extended the environment, there grew the need to add other agents that were able to help the visitor. This raised our interest in having these agents communicate with each other as well and to endow them with some form of autonomous behavior. Rather than towards a goal-directed information and transaction system comparable to 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 show a wide variety including assigning personalities and emotions to artificial agents, usability studies involving a navigational assistant, formal specification of (interactions in) virtual environments and reinforcement learning for agents in this multimodal environment to increase their autonomy.

When we enter our Virtual Muziek Centrum, we see the information agent called Karin, waiting to tell us about performances, artists and available tickets. Visitors can explore this virtual environment, walking from one location to another, looking at posters, clicking on objects and so on. Karin can be asked natural language questions about performances in the theater. She has access to a database containing all the performances in the various theaters during the current season. Karin has a 3-D face that allows simple facial expressions and simple lip movements that are synchronized with a text-to-speech system mouthing the system's utterances to the user (see Nijholt & Hulstijn [9] for details). Other agents have been introduced in this environment. For example, a navigation agent, that knows about the geography of the building and that can be addressed using typed in natural language utterances. The visitor can ask the agent about existing locations in the theater. When the request is understood, a route is computed and the viewpoint in the world is guided along this route to the destination. The navigation agent has not been visualized as a 3D embodied agent.

A Java based agent framework has been introduced to provide the protocol for communication between agents and the introduction of other agents. For example, why not allow the visitor to talk to the map of the seats in the main concert hall or to a poster displaying an interesting performance? In fact, we can have a multitude of potential and useful agents in our environment, where some just trigger an animation, others can walk around and others have built-in intelligence that allows them to execute certain actions based on interactions with visitors. Some of the 3D avatars that live in our environment have not yet been incorporated in the frame-

work in a way that visitors can communicate with them (a baroque dancer, a piano player). We have been experimenting with embedding our environment in a multi-user shell (Reitmayr et al. [11]) that allows to entertain multiple visitors that can make themselves visible to each other as avatars (VRML objects). These avatars move along with the visitor, but they can also be assigned animations, intelligence and interaction abilities. Hence, we can have different human-like agents. Some of them are autonomous embodied agents standing or moving around in the virtual world and allowing interaction with visitors of the environment. Others represent human visitors of the environment. We want any visitor to be able to communicate with autonomous agents and visitors, whether visualized or not. That means we can have interactions between agents, between visitors, and between visitors and agents. This is a rather ambitious goal which cannot be realized yet completely.

3.4 Interacting about Performances and Environment

How does interaction between domain agents and visitors take place? We decided to introduce a model of natural language interaction between Karin and user that is rather primitive from a linguistic point of view, but sufficiently intelligent from a practical and pragmatic point of view. This natural language understanding system mediates between the user and a database containing information about performances, artists and prices. Although the ‘linguistic intelligence’ is rather poor, the outcome of a linguistic analysis can be passed on to pragmatic modules that produce relevant system responses in the majority of cases. The system prompts make users adapt their behavior to the system. Karin presents her information using text-to-speech synthesis and lip movements. When there are too many performances to read out, she presents a table and draws the user’s attention to this table using eye movement and a natural language utterance. The dialogue system can interpret and generate references to items in this table.

It may be clear how to address Karin. However, visitors may want to address other domain agents and agents that represent users. As mentioned, this is work in progress. We are following several approaches to solve this problem. They are related and can be integrated since all of them are agent-oriented and based on a common framework of communicating agents. In addition, we have built this framework in such a way that different agents with different abilities can become part of it: a simple animated piano player, a baroque dancer that ‘understands’ the music she is dancing on, Karin who knows about theater performances, and a navigation agent that knows about the geography of the building.

Developing navigation agents leads to a number of questions. How can we build navigation intelligence into an agent? What does navigation intelligence mean? How can we connect this intelligence to language and vision intelligence? Visitors of our environment are language users and, moreover, they know and interpret what they see. There is a continuous interaction between verbal and non-verbal information when interpreting a situation in our virtual environment. This interaction and the representation and interpretation of sources and then the generation of multimedia from them are among the main topics of our research.

We very much follow Darken & Silbert [3] in our approach to navigation. To assist the visitor in navigating through our virtual theater, we have added both a map and an intelligent navigation agent. The visitor can ask questions, give commands and provide information when prompted by the agent. This is done by typing natural language utterances or by moving the mouse pointer over the map to locations and objects the user is interested in. On the map the user can find the performance halls, the lounges and bars, selling points, information desks and other interesting locations and objects. The current position of the visitor in the virtual environment is marked on the map. While moving in VR the visitor can check his or her position on this map. When using the mouse to point at a position on the map, references can be made by both user (in natural language) and system to the object or location pointed at. We have annotated a small corpus of example utterances that appear in navigation dialogues. An example of a question is: “What is this?” while pointing at an object on the map, or “Is there an entrance for wheel chairs?”. Examples of commands are “Bring me there.” or “Bring me to the information desk.” Examples of short phrases are “No, that one.” or “Karin.” From the annotated corpus a grammar was induced and our unification-type parser for Dutch can be used to parse these utterances into feature structures. Three agents communicate to fill in missing information in the feature structure and to determine the action that has to be undertaken (answering the question, prompting for clarification or missing information, displaying a route on the map or guiding the user in VR to a certain position). The navigation agent, the dialogue manager and the Cosmo Agent do this in co-operation. Not yet implemented is the possibility that not only the position but also what is in the eyesight of the visitor is known. This will allow interpretation of references to objects that are visible to a visitor.

3.5 Towards a Theater Community

The length of this paper does not allow a comprehensive survey of all the problems we have to deal with when we want an agent-oriented design of our environment and have it inhabited by agents that can be embodied, have intelligence and personality and can communicate with each other and with agents that represent visitors. To design and maintain an environment like that we need some uniformity from which we can diverge in several directions: agent intelligence, agent interaction capabilities, agent visualization and agent animation (cf. Nijholt & Hondorp [10]). Standards are needed to allow frameworks for communication, internal modelling, and animation of embodied agents. These standards should also address issues concerned with multi-user and multi-developer environments.

In Egges et al. [4] we introduce an approach to the internal modelling of agents we think we can use in our multi-agent and multi-user environment. Our approach discussed there, is limited, but nevertheless allows modeling of ‘intelligence’ in terms of beliefs, desires and plans, and possible extensions to the modeling of emotions and an agent’s knowledge about movements, postures and non-verbal communication. Our current emotion research is reported in Kesteren et al. [6] and Bui et al. [2].

References

- 3.1 J. Cassell, J. Sullivan, S. Prevost & E. Churchill (eds.). *Embodied Conversational Agents*. MIT Press, Cambridge, 2000.
- 3.2 Bui The Duy, D. Heylen, M. Poel & A. Nijholt. Generation of facial expression from emotion using a fuzzy rule based system. Submitted for publication, July 2001.
- 3.3 R.P. Darken & J.L. Silbert. Way finding strategies and behaviors in virtual worlds. *CHI'96*, 142-149.
- 3.4 A. Egges, A. Nijholt & R. op den Akker. Dialogs with BDP Agents in Virtual Environments. In: Proceedings 2nd IJCAI workshop on *Knowledge and Reasoning in Practical Dialogue Systems*. Seattle, Washington, August 2001, to appear
- 3.5 W. L. Johnson, J. W. Rickel, and J. C. Lester. Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments. *The International Journal of Artificial Intelligence in Education* (2000) 11, 47-78.
- 3.6 A.-J. van Kesteren, R. op den Akker, M. Poel & A. Nijholt. Simulation of emotions of agents in virtual environments using neural networks. In: *Learning to Behave: Internalising Knowledge*. Proceedings Twente Workshops on Language Technology 18 (TWLT 18), joint Cele-Twente workshop, November 2000, 137-147.
- 3.7 J.C. Lester, S. G. Towns, C.B. Callaway, J.L. Voerman & P.J. Fitzgerald. Deictic and emotive communication in animated pedagogical agents. In: Cassell et al., 2000.
- 3.8 A.B. Loyall. *Believable Agents: Building interactive Personalities*. PH.D. Thesis, CMU-CS-97-123, May 1997, Carnegie Mellon University.
- 3.9 A. Nijholt & 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.
- 3.10 A. Nijholt & H. Hondorp. Towards communicating agents and avatars in virtual worlds. In: Proc. *EUROGRAPHICS 2000*, A. de Sousa & J.C. Torres (eds.), August 2000, Interlaken, 91-95.
- 3.11 G. Reitmayr et al. Deep Matrix: An open technology based virtual environment system. *The Visual Computer Journal* 15: 395-412, 1999.
- 3.12 K. Takahashi, J. Kurumisawa & T. Yotsukura. Networked theater. Proc. First IEEE *Pacific-Rim Conf. On Multimedia*, December 2000, University of Sydney, Australia.
- 3.13 N. Tosa & R. Nakatsu. Emotion recognition-based interactive theater – Romeo & Juliet in Hades. Proc. *EUROGRAPHICS'99*, M. Alberti et al. (eds.), 179-182.
- 3.14 H.H. Vilhjalmsson. Autonomous communicative behaviors in avatars. Master's Thesis, MIT Media Laboratory, 1997.
- 3.15 T. Watanabe, M. Okubo & Y. Ishii. An embodied virtual face-to-face communication system with virtual actor and virtual wave for human interaction sharing. In: Proc. 4th World Multiconference on *Systemics, Cybernetics and Informatics (SCI'00)*, Vol. III, Orlando, USA, 146-151, 2000.