Embodied Agents in Virtual Environments

The AVEIRO Project

Dirk Heylen, Anton Nijholt and Mannes Poel
Computer Science Department – University of Twente
PO Box 217 – 7500 AE Enschede
Phone: +31-53-4893740, Fax: +31-53-4893503
Email: {heylen,anijholt,mpoel}@cs.utwente.nl

ABSTRACT: We present current and envisaged work on the AVEIRO project of our research group concerning virtual environments inhabited by autonomous embodied agents. These environments are being built for researching issues in human-computer interactions and intelligent agent applications. We describe the various strands of research and development that we are focussing on. The undertaking involves the collaborative effort of researchers from different disciplines.

KEYWORDS: agents, virtual environments, virtual humans, human-computer interaction, intelligent tutoring systems

INTRODUCTION

The AVEIRO project of the Parlevink Research group at the Computer Science Department of the University of Twente, groups together different research strands that are involved in constructing a virtual environment inhabited by intelligent agents. To study various issues in the field of human-computer interaction (including multi-modality, speech and natural language understanding, affective computing) and intelligent agent technology, we have created a virtual replica of one of the local theatres (the Virtual Music Centre, VMC). The agents are used to perform specific tasks (tutoring, reception, navigation). Some of them are embodied and capable of conversing face-to-face with visitors of the environment. In a multi-user version of this environment visitors are also represented by embodied agents (with very limited capabilities).

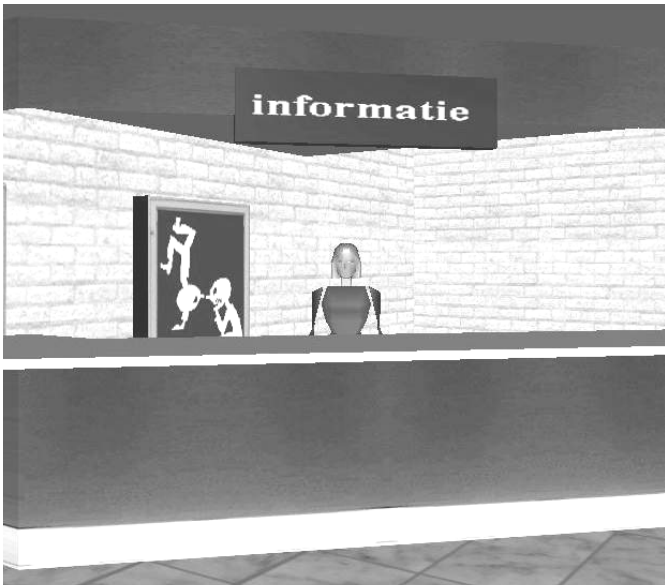

Figure 1: Karin at the information desk of the VMC

The environment has been built using VRML (Virtual Reality Modeling Language) and is accessible on WWW[1]. Originally the environment was built around an existing natural language dialogue system that is dedicated to provide information about theatre performances and that allows reservations for these performances to be made ([9]). In the virtual environment, the dialogue system has been assigned to a visualized embodied agent (Karin), cf. Figure 1.

Once we had created 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 ([13]).

[1] http://parlevink.cs.utwente.nl.

RESEARCH ISSUES

Our research group is made up of researchers with quite divergent interests and backgrounds. Domains of study vary from graphics to speech and natural language processing, from neural networks and learning to software engineering and formal specification languages.

Artificial Intelligence/Intelligent Agents One of our major concerns is the study of various topics in the area of artificial intelligence. Our virtual environment is inhabited by more or less intelligent agents that can perform specific tasks. The agents are used to research different aspects of intelligence (reasoning, language understanding, learning), using a variety of models of intelligence, including neural networks as well as symbolic models.

Human-Computer Interaction The virtual environment was initially constructed around a previously developed natural language dialogue system. Using this system, users could get information about performances and buy tickets by telephone. The receptionist Karin in the virtual environment is an embodied agent that incorporates this dialogue system. The virtual setting, allows us to explore other issues in human computer interaction, as well, such as multi-modal interaction, the use of different input and output devices typical for VR, and the combination of both verbal (natural language) and non-verbal communication (gestures, facial expressions, gaze).

Human factors Our interest in human-computer interaction is not limited to researching how to design and build a virtual environment, but also extends to investigating and evaluating the use that can and is being made from these systems, their effects and efficiency. Human factors is therefore another approach that is being paid attention to in our group.

Graphics The higher level functionality of the system has to be put to work in a graphical system that dynamically visualises the objects, the agents, and the environment. Although we are relying mostly on third-party software to build and display the graphical environment, we are also building up expertise to allow us to tune the systems to our specific needs and integrate the graphical level with the higher levels.

Software engineering The original environment was constructed, more or less from scratch. In the meantime several versions of the environment have been created to experiment with a specific problem. In the AVEIRO project we want to engineer the environment again, using an agent based software engineering approach, paying special attention to modularity, specification, verification and maintainability issues.

In the following sections we provide an overview of the various research and development action lines that make up the AVEIRO project.

AGENTS: MIND AND BODY ARCHITECTURE - BEHAVIORS

The virtual environment is inhabited by smart objects and intelligent agents, each fulfilling a specific task. Some of these agents are embodied and capable of natural language interactions with visitors. These embodied conversational agents are used to supply information or for tutoring, transaction and negotiation. Different aspects of artificial intelligence research are carried out by experimenting with different implementations of these agents. We are working on an agent platform that allows a common definition style for different agents and interfaces for communication and interoperability.

MINDS

Different models of intelligence are investigated within the AVEIRO project. The approach to natural language dialogues with the agents, is embedded in a symbolic tradition. The models of conversational action are tied more closely to BDI models of rational action. In other areas and for other tasks like learning, we also experiment with subsymbolic models, using neural networks and other techniques. The minds of our agents not only incorporate models of intelligence and rational action, but are complemented with theories of emotion and affect. The conceptual state of the agent is constantly updated by taking into account changes that the agent senses.
Among the topics of research we are investigating, we point out the following.

Mental Models
In [4], a logical approach to modelling the mental state of conversational agents was investigated. A BDP (belief, desire, plan) model was introduced and a logical language (representing beliefs, the state of the world as well as the meaning of utterances) was defined to specify the mental state of an agent. A reasoning mechanism is provided that allows the agent to execute its plans. The system serves as a research and specification platform in which the symbolic architecture of rational planning and decision making can model certain components of the agents tasks, like, for instance the dialogue management module.

Some of our agents are conceived as embodied human-like agents. In order to build believable agents it is not sufficient to endow them with intelligence that enables them to solve complex problems. They should also be provided with a personality, moods and emotions ([3],[17]). In [7], we present a hybrid architecture, including neural networks, for a cognitive model of agents in which emotions play a role. The system implements an appraisal model of emotions based on the theory proposed by Ortony, Clore and Collins ([15]). A gridworld has been created in which objects and other agents live and different events may occur. The agents have some knowledge about the world, the events that can occur (it might rain, grass may grow, pools can be filled with water, apples may grow on trees, some agents are predators, others are preys), and the probabilities of these events. They have a physical condition. Moving around in the world, they try to survive: eating when they get hungry and food is in sight, trying to fend of predators, etcetera. Their cognitive system includes a representation of their emotional state. Neural networks are used to update the emotional state based on the previous state and the events occurring in the world (they act as event-appraisers). We hope to be able to use this model and architecture for the embodied conversational agents that live in the Virtual Theatre Environment and link the emotional state to expressive behaviours like facial expressions (see the section facial expressions).

Learning
The agents in the gridworld are trained to associate emotions with events. Learning agents are an effective way of tackling many problems in complex, dynamic, multi-agent environments. We are investigating the use of *reinforcement learning*. In particular, we are looking at *constructive* neural networks and learning of higher order concepts as a way to put more autonomy in the agents, making design of reinforcement learning structures less ad hoc and making them more adaptable to dynamic, user-inhabited environments and multi-agent environments.

BODIES

The agents live in a virtual environment. The environment and some of the agents have visual representations in the graphical world. We use VRML as the general representation language for the three dimensional world and use Java as the general programming language. We are currently also experimenting with Java 3D with a view to exporting the world to a new language, perhaps in the near future. Given that we want the world to be accessible via the internet, we must keep the graphical complexity low. This is one of the reasons to choose cartoon-faces for the agents. For some of our experiments, we use graphical representations that are more complex. Those are not part of our basic theatre environment.

For the graphical part we rely on existing software as much as possible and comply with standards wherever possible. However, we are also developing some tools that should make it easier to develop embodied agents in simple VRML worlds. One of the systems that is being developed will be composed of a mesh editor for modeling the shape of the agents, a movement editor for basic movement types, a motion control module to concatenate basic movement and a a behavior control module to perform tasks in response to the received events. The systems are specifically designed to be simple to use and deliver objects with low graphical complexity as required by the constraints on current bandwidth ([8]).

BEHAVIOUR – TASKS

For the execution of the specific tasks the agents should also be capable of some common, general behaviors. They should be able to perceive the world and the users and to interact with them: manipulate objects, move around in the world and, in the case of embodied conversational agents, have a *face-to-face conversation* with the user. Each agent has a specific task in the environment: providing the user with information on performances and selling tickets, helping the user *navigate* through the environment and providing information about the location, or *tutoring*.
Providing these different functionalities in different agents allows us to experiment with different approaches to similar problems or to research specific issues in more depth. We will present some details about our approach to the general conversational behaviour and two specific tasks: tutoring and navigation/assistance.

Conversation

The study of embodied conversational agents or virtual humans is one of the main reasons for investigating agents and placing them in a virtual environment. The original dialogue system that was developed prior to the construction of the virtual environments is now re-used for the dialogue capabilities of Karin, the receptionist agent.The embodiment makes it possible and necessary to investigate other modes of communication as well. Several projects are looking therefore at some issues in non-verbal communication (facial expressions, gestures, gaze).

The sequence of utterances, consisting of alternating turns by the interlocutors, is governed by different types of constraints that must be taken into account for the dialogue to be felicitous. Utterances must be appropriate to the context. Proper human-like agents should know when it is their turn to speak, what conversational action their utterance should involve, what information it should convey (in accordance with the overall goal of the conversation and the particular stage) and in what modality or combination of modalities it is expressed best. They should be able to formulate their utterance correctly, accompanied by the proper facial expressions, gestures, gaze, and body posture. While speaking they must monitor the hearer, seeing whether he is paying attention, looking for expressions of puzzlement or interest, and listening for backchannel vocalisations. When the visitor is talking, the agent should assume the appropriate listening posture, listen carefully, and decode what the speaker is trying to get across. In this respect building embodied conversational agents is an excercise in affective computing.

Our basic version of the dialogue system involves a simple rewrite (pattern-matching) system to deal with natural language analysis ([9]). The model is adaptable to the specific topic- and goal-restricted conversations each agent is assumed to be able to engage in. Our current work involves making the natural language analysis and generation components more sophisticated using constraint-based grammars ([1]), embedding the dialogue and conversational actions in a general BDP model of rational action (see above), dealing with all kinds of uncertainties that arise in the dialogue situation using bayesian networks ([6]), and integrating more sophisticated non-verbal communication behaviors. Among the latter, we are looking at gaze and facial expressions, at the moment.

Gaze Using eye-trackers, it is possible to monitor which agent or part of the environment the user is looking at. This may trigger certain actions from the agents. For instance, in a face-to-face conversation, (mutual) gaze contributes to regulating the flow of conversation and more specifically turn-taking behaviour. We are investigating to what extent the implementation of similar behavior in agents improves conversational interactions ([12],[19]).

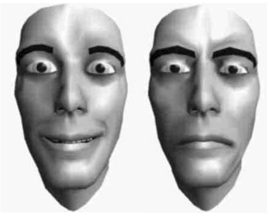

Figure 2: A happy and a sad face

Facial expressions The face provides all kinds of important clues during a conversation. We are trying to provide our embodied conversational agents with the correct and appropriate expressive behaviour. Facial expressions should be in tune with the state of the conversation (content, context), the cognitive state of the agent (does it understand what is being said, does it believe what is being said, does it enjoy itself) and the interactional parameters (show the wish to interrupt, listen attentively). We have particularly, looked at linking the emotional state of the agent to appropriate expressions ([2]). In a prototype (figure 2) extending the expressive face of ([16]). an emotional state is transformed, using a fuzzy logic rule based system, to a set of muscle contractions.These muscle contractions determine the facial expression. The fine tuning of the fuzzy rule based system is done by hand.

Tutoring

Several of our agents are involved in tutoring, including a piano teacher ([11]). This is a common task of embodied conversational agents. It allows us to research specific aspects of interaction between agents, visitors and the virtual environment. We are aiming at implementing general instruction models that are independent of specific application domains.

Jacob ([5]) is a prototype of a virtual instruction environment. The agent Jacob provides instruction and assistance for the Tower of Hanoi problem, cf. figure 3. The user can move the blocks in the virtual environment and Jacob will monitor this performance, judging the correctness of the procedure, possibly giving advice or suggesting corrections, or illustrating a manoeuvre by doing it himself. The main objectives in building this

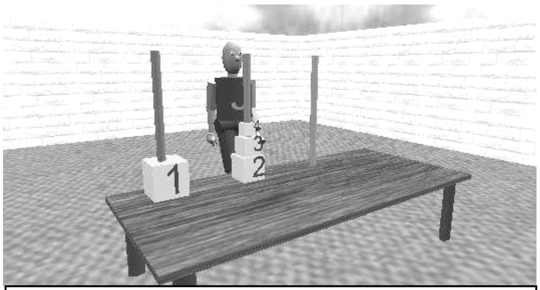

Figure 3: The instruction agent Jacob

agent were (i) to research general adaptable task and instruction models that make up the core of the instructor's mind, (ii) to look more closely at multi-modal interactions where users and agents induce changes in the environment that require continuous updates and processing by the agent, (iii) to investigate user models as tutors have to be able to reason about the rationale behind the user's moves.

Navigation and Assistance

The Karin agent knows about the performances in the theatre. Questions about the environment itself can be answered by the navigational assistant who can also help the user to navigate through the building ([10]). The navigation agent can show the current position of the visitor on the map (for which it has to communicate with the browser agent to locate the user in the VRML world). Understanding user requests (and providing the answers) involves a lot of integration of information from different sources. The context for reference resolution involves the current, past and future positions of the visitor. Also the user can make references to what he sees in the environment or on the map. The assistant can point out places on the map or provide language descriptions, it can take the user where he wants to go, or indicate a route to follow on the map ([11]).

SOFTWARE ENGINEERING ISSUES

The work on the agents and the environment is in a continuous flux. Not only do we want to extend the environment and the agents with new functionalities but we also want to improve upon existing modules or experiment with a variety of techniques and solutions. Different working versions of the environment exist in parallel at the same time, while work on some isolated issue is explored in depth. If we think the work provides interesting features to incorporate in a next demo-version, we try to do so.
The original environment was not set up with this type of modularity and extendibility requirements in mind. Although this has not led to insurmountable problems so far, we think it is interesting to look at (agent based) software engineering methods that will lead to more flexible architectures that allow for plug-and-play functionality ([20]).

As a corrolary of these development concerns, we are also looking at other software engineering issues surrounding the construction of complex virtual environments inhabited by multiple interacting agents. Specifically we are concerned with what kind of specification languages are best used to model distributed virtual environments (VEs) with environment-aware dialogue agents ([18]). In many models, a system is described by means of interacting objects or processes. The manner of interaction, however, differs. For example, there is a variety of message passing mechanisms, ranging from synchronous to fully asynchronous, and a variety of data sharing mechanisms, such as shared variables, and blackboard, publisher-subscriber, and dataflow models. Furthermore, there are different ways in which the behaviour of an individual process may be specified, for example, as an event loop or production rule system, or as a state automaton. A software toolkit with corresponding specification language is being developed, based upon a choice of existing concepts from the literature. It serves as a testbed for the applicability of these concepts for VE development.

EVALUATION AND HUMAN FACTORS

Computer models of simulated humans in virtual worlds can be evaluated in terms of their computational properties: simplicity of the models, efficiency (resources, speed, fluency), and the extent to which they implement the necessary functions correctly and completely. However, when dealing with interface systems it is important to evaluate how well they work for users interacting with the system, as well.

A technological new project, like our virtual music center, is in the initial phase always technology pushed. In an experimental setting the possibilities and bottlenecks of the new technology are tested, without looking at the usability for, and wishes of, the potential users of the system. This project aims at the investigation of the usability, for the prospective users or user groups, the behavior of users in such a virtual environment, their wishes and preferences. The ultimate goal will be to develop a user model that can be used for the development or adoption of the (to be) developed intelligent agent based virtual environment.

CONCLUSION

Research and development of virtual environments inhabited by multiple agents, interacting with each other and the visitors of the environment using natural language and other modalities, is a complex undertaking involving different types of expertise. This makes it a particularly rich and inspiring project to work on together as a group consisting of computer scientists, mathematicians, logicians, linguists, psychologists and researchers in cognitive ergonomics.

REFERENCES

[1] Akker, R. op den, A. Nijholt (2000). Dialogues with Embodied Agents in Virtual Environments. In: Proceedings 2nd International Conference on Natural Language Processing: NLP 2000: Filling the gap between theory and practice. D.N. Christodoulakis (ed.), Lecture Notes in Artificial Intelligence 1835, Springer, p. 358-369.

[2] Bui, T.D, D. Heylen, M. Poel, A. Nijholt (2001) Generation of facial expressions from Emotion using a Fuzzy Rule Based System, accepted for the The 14th Australian Joint Conference on Artificial Intelligence

[3] Cassell, J., J. Sullivan, S. Prevost, E. Churchill (eds) (2000). Embodied Conversational Agents. MIT.

[4] Egges, A. (2001). Conversational Agents in a Virtual Environment with Multi-modal Functionality. Masters Thesis, Computer Science Department, Enschede.

[5] Evers, M., A. Nijholt (2000). Jacob – An animated instruction agent in virtual reality. In: Proceedings 3rd International Conference on Multimodal Interfaces (ICMI 2000). Lecture Notes in Computer Science. Beijing, p. 526-533.

[6] Keizer, S. (2001). A Bayesian Approach to Dialogue Act Classification. In: P. Kühnlein, H. Rieser, H. Zeevat (eds.) Proceedings BI-Dialog 2001. 5th Workshop on Formal Semantics and Pragmatics of Dialogue, Bielefeld.

[7] Kesteren, A-J., R. op den Akker, M. Poel, A. Nijholt (2000). Simulation of Emotions of Agents in Virtual Environments using neural networks. In: Learning to Behave: Internalising Knowledge. Proceedings of the Twente Workshop on Language Technology 18, Enschede, p. 137-147.

[8] Kiss, Sz. (2001). Web Based VRML Modelling. In Proceedings 5th International Conference on Information Visualisation, IV2001, London.

[9] Lie, D., J. Hulstijn, A. Nijholt, R. op den Akker (1998). A Transformational Approach to NL Understanding in Dialogue Systems. Proceedings NLP and Industrial Applications, Moncton, New Brunswick, p. 163-168.

[10] Luin, J. van, R. op den Akker, A. Nijholt (2001). A Dialogue Agent for Navigation Support in Virtual Reality. In: J. Jacko and A. Sears (eds.) Proceedings CHI 2001: Anyone Anywhere. ACM, p. 117-118.

[11] Nijholt, A. (2001). Agent Assistance: From Problem Solving to Music Teaching. In: M. Beer, J. Whatley (eds.) Proceedings Agents and Internet Learning. Workshop at 5th International Conference on Autonomous Agents, Montreal, p. 11-13.

[12] Nijholt, A., D. Heylen, R. Vertegaal (2001). Inhabited Interfaces: Attentive Conversational Agents that help. In: P. Sharkey (ed.) Cyberpsychology and Behavior, Vol. 4, Special Issue (to appear).

[13] Nijholt, A., J. Hulstijn (2000). Multimodal Interactions with Agents in Virtual Worlds. In: N. Kasabov (ed.), Future Directions for Intelligent Systems and Information Science, Studies in fuzziness and soft computing. Physica-Verlag, p. 148-173.

[14] Nijholt, A., J. Zwiers, B. van Dijk (2001) Maps, Agents and Dialogue for Exploring a Virtual World. In: Proceedings of the 5th World Multiconference on Systemics, Cybernetics and Informatics (SCI 2001). Orlando, p. 94-99.

[15] Ortony, A., G. Clore, A. Collins (1988). The Cognitive Structure of Emotions. CUP.

[16] Parke, F.I. and K. Waters (1994). Computer Facial Animation. AK Peters.

[17] Picard, R. (1997). Affective Computing. MIT Press.

[18] Schooten, B. van (2000). A Specification Technique for Building Interface Agents in a Web Environment. In: A. Nijholt, D. Heylen, K. Jokinen (eds.) Learning to Behave. TWLT 17. Workshop on Interacting Agents, Enschede, p. 83-98.

[19] Vertegaal, R., R. Slagter, G. van der Veer, A. Nijholt (2001). Eye Gaze Patterns in Conversations: There is more to conversational agents than meets the eyes. In: J. Jacko, A. Sears, M. Beaudouin-Lafon, R.J.K. Jacob (eds.) Proceedings CHI 2001: Anyone Anywhere, p. 301-308.

[20] Zwiers, J., B. van Dijk, A. Nijholt, R. op den Akker (2000). Design Issues for Intelligent Navigation and Assistance Agents in Virtual Environments. In: A. Nijholt, D. Heylen, K. Jokinen (eds.) Learning to Behave. TWLT 17. Workshop on Interacting Agents, Enschede, p. 119-132.