Fourth International Conference on knowledge-Based Intelligent Engineering Systems & Allied Technologies, 30th Aug-1st Sept 2000, Brighton,UK

Agent and User Inhabited Virtual Communities: A Case Study

Anton Nijholt (Parlevink Research Group)[1]
University of Twente, PO Box 217
7500 AE Enschede, the Netherlands
anijholt@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. This environment is a laboratory for research and experiments on 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. Although the environment is tuned to a theatre environment, we think there are sufficient general properties in order to learn about other applications, e.g. other theme oriented, educational and entertainment environments and even electronic commerce environments. In addition, especially in the home environment, there will be a growing need for social interfaces that are inhabited by visualized domain agents, user agents, friends and relatives that help, advise, and 'negotiate' on matters that range from what to prepare for dinner to how to end a relationship.*

1 Introduction

In [2] we discussed a natural language dialogue system that offered information about performances in some of our local theatres and that allowed visitors to make reservations for these performances The intelligence of this system showed in the pragmatic handling of user utterances in a dialogue. The 'linguistic intelligence' was rather poor, however the outcome of a linguistic analysis could be given to pragmatic modules which in the majority of cases (assuming 'reasonable' user behavior) could produce system responses that generated acceptable utterances for the user. With this we don't mean that for any user utterance the next system utterance could be considered as a satisfactory answer or comment. Rather it should be considered as an utterance containing cues how to continue the dialogue in order to come closer to a satisfactory answer. The general idea behind the system was that users learn how to phrase their questions in such a way that the system produces informative answers. The system prompts can be designed in such a way that users adapt their behavior to the system and the prosody of system utterances (in a spoken dialogue) can invite user's to provide information that they (incorrectly) already assumed to be known by the system, making the interaction more natural.

2 Building a Virtual Theatre

We embedded this NL accessible theatre information and booking system in a virtual reality environment that allowed visitors to walk around in the theatre and to go to an information desk. The theatre was built according to design drawings of the architects of the building. Visitors can explore this environment, walk from one location to another, ask questions to available agents, click on objects, etc. Karin (see Fig.1), 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 [3].

Other agents in this environment have been introduced. One example is a navigation agent which knows about 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 and 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

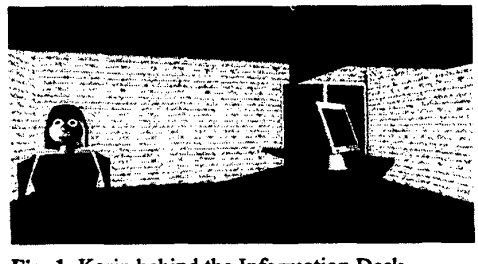

Fig. 1 Karin behind the Information Desk

[1] 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.

0-7803-6400-7/00/$10.00 ©2000 IEEE

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 on 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.

Fig. 2 Visitor (Jacob), Baroque Dancer and Piano Player

3 Towards a Theatre Community

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.

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 agents. For example, we have a piano player on stage with some simple predefined animations, there is a baroque dancer (imported from the Baroque Dance Project [1]) with animations synchronized with audio and there are visitors, able to move around in the multi-user environment. In Fig. 2 we see a visitor's avatar that has been so impertinent to climb the stage in order to get a closer look at the performing dancer. Its animations allow it to walk around following the coordinates of the moving viewpoint position of its owner.

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.

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 [7]. 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.
- Living Worlds (LW) Standard [8]. The aim is to define a conceptual framework and specify interfaces to support the creation of multi-user and multi-developer applications in VRML [8]. 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.

The visitor's avatar shown in Fig. 2 has been built following the H-anim standard. Presently we use the DeepMatrix [4] 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 LW specufication are the issues which will make it have to be dealt with in order to introduce standardized interacting agent frameworks in virtual environments. In conclusion, we think that for our environ-

ment 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.

4 Problems in Interaction Modeling

There exist many linguistic and dialogue modeling problems that are specific for multiple dialogue partners present in a virtual environment that have hardly been investigated in natural language processing research. We shortly address some of these questions for agents in our virtual environment:

How does an agent know that the user addresses a message to him?

For example, the visitor can either refer to an agent by naming (using a definite address or proper name) it or by gazing at him accompanied by some natural language indicator (like "you" in "can you tell me how to get to the main hall ?"). Agents should know how they can be addressed.

How does the agent know the users intention?

Sometimes syntactic/lexical clues are sufficient to identify the conversational act a user performs by uttering a sentence. For example the sentence: "what color does this box have?" gives sufficient syntactic/lexical information to identify the user's wh-question after some particular attribute of a particular object that is explicitly mentioned in the utterance. Often we need dialogue information, knowledge of the application (domain) and knowledge about the user to identify the intention of the user. Therefore conversational agents need knowledge of users, and application domain to identify the intention of the acts performed by the user during a conversation.

Fig. 3 Conversation with Attentive Agents

How does the agent know the referents of the natural language expressions the user has uttered?

The agent has a language model of the communication language and obtains a semantical representation from a parser. The semantic representations output by the parser are interpreted in the context of the dialogue and in the context of the virtual environment. In order to find the denotations of objects in the environment these objects (or rather their abstract counterparts as objects) have labels that are natural language indicators. When a user points (for instance by a mouse pointer, or gazing) to the graphical representation of some object (a chair say) in the virtual environment and asks the agent to move the object; the object will be put in focus of the dialogue. The interpretation function of the semantic representation of the input sentence "move this chair to the main hall", will look for some object in the dialogue focus that matches the word "chair". Likewise the action name "move" gets a denotation from the action names of the agent. If the agent is ready to perform the action denoted by "move" it will perform it on the object denoted by "this chair". Another solution would be that the agent seeing the word "chair" has to find some object that matches his image of a "chair" by means of pattern-matching of the "geometrical" representation of the object.

To gain experience with the gaze modality, we are implementing findings on gaze [6] in a separate prototype environment with an eyetracking system that establishes where a user looks at. Muscle models are used for generating accurate 3D facial expressions. The goal is that in an environment with more than one agent, each agent is capable of detecting whether the user is looking at him, and of combining this information with speech data to determine when to speak or listen to the user. To help the user regulate conversations, agents generate gaze behavior as well. This is exemplified by Fig. 3. The agent on the left is the focal point of the user's eye fixations. The right agent observes that the user is looking at the speaker, and signals it does not wish to interrupt by looking at the left agent, rather than the user. In experiments we have a set-up with a user and two agents where the agents have related tasks. In the experiment we make a more explicit distinction between the information task and the reservation task of our information and transaction agent Karin. We use a Karin_1 and a Karin_2 who have to communicate with each other (exchange information

about user and chosen performance) and with the visitor. When in the reservation phase with Karin_2 it turns out that the desired number of tickets is not available or that they are too expensive, it is necessary to go back to Karin_1 in order to determine an other performance.

5 Problems in Knowledge & Software Engineering

As mentioned in section 3, a framework that allows uniformly modeled agents (with different levels of intelligence) is needed. Appearance, behavior and intelligence can be task and domain dependent. When an agent in the framework does not have the knowledge to come up with a sufficiently adequate act when addressed by a user, there is of course the possibility to start a dialogue with the goal to get more information, but it may also be possible to delegate a task to an other agent. As a simple example, when Karin does not understand the question or can not find an answer in the database we can try an 'Ask Jeeves' approach on the Web. A search on the Web can be supported by an ontology of the domain and domain reasoning. Clearly, when a user says something like: "Well, I forgot the name of the actress, but I know she's married to Tom Cruise.", then it is likely that the name can be retrieved from WWW and be used to fill in the missing part of a user's question. In this particular case, using the Northern Light search engine, we got 1497 hits, where the first hit was a profile of Tom Cruise containing the following information:

> Wife: Mimi Rogers. Actress. Married May, 1987. Divorced January 1990. Nicole Kidman. Actress. Married December 24, 1990.

Lower in the list of hits we find actress Meg Ryan:

> ". . . but she gained good notice for her next assignment, a solid supporting turn in the jingoistic Tom Cruise actioner Top Gun (1986), in which she was cast as the wife of Cruise's naval fighter co-pilot, played by . . . "

Hence, although the information is available we need linguistics and common sense modeling (a divorce overrules a marriage, it is the last marriage that counts, . . .) to make this approach effective.

The second type of problems we want to mention are the software engineering problems. Virtual environments may feature a variety of interactive objects, agents which may use natural language to communicate, and multiple simultaneous users. All may operate in parallel, and may interact with each other concurrently. Next to this, the possibility of using virtual reality techniques to enhance the experience of virtual worlds offers new ways of interaction, such as 3D navigation and visualization, sound effects, and speech input and output, possibly used so as to complement each other.

One line of research we have taken is an attempt to address these issues by means of a formal modeling technique that is based on the process algebra CSP (Schooten [5]). A simplified flow of interaction has been specified, showing all relevant interaction options for any given point in time. The system architecture has been modeled in an agent-oriented way, representing all system- and user-controlled objects, and even the users themselves, as parallel processes. The interaction between processes is modeled by signals passing through specific channels. Interaction modalities (e.g., video versus audio and text versus graphics) may also be modeled as separate channels.

This modeling technique, which will be elaborated in the future, has some strong points. It enables a clear and unambiguous specification of system architecture and dynamics. It may be useful as a conceptual model, modeling the fact that a user experiences interaction with other users and agents in a similar way than in a completed system. And finally, it enables automatic prototyping, such as architecture visualization and verification of some system properties.

Acknowledgements: I'm grateful to Rieks op den Akker for his contribution to section 4 of this paper.

References

[1] M. Bertolo, P. Maninetti & D. Marini. Baroque dance animation with virtual dancers. *Eurographics '99*, Short Papers and Demos, Milan, 1999, 117-120.

[2] D. Lie, J. Hulstijn, R. op den Akker & A. Nijholt. A Transformational Approach to NL Understanding in Dialogue Systems. Proc. *NLP and Industrial Applications*, Moncton, 1998, 163-168.

[3] 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.

[4] G. Reitmayr et al. Deep Matrix: An open technology based virtual environment system. *The Visual Computer Journal*, to appear.

[5] B. van Schooten et al. Modeling Interaction in Virtual Environments using Process-algebra. Proc. *Interactions in Virtual Worlds*, 1999, University Twente, 195-212.

[6] R. Vertegaal et al. Why conversational agents should catch the eye. Proceedings *CHI 2000*, 2000.

[7] VRML Humanoid Animation Working Group, http://ece.uwaterloo.ca/~h-anim/, 1998.

[8] VRML Living Worlds Working Group: Making VRML 97 Applications Interpersonal and Interoperable. http://www.vrml.org/living-worlds, 1998.