An Agent-based Virtual Theatre Community

Anton Nijholt
Centre for Telematics and Information Technology (CTIT)
University of Twente, PO Box 217
7500 AE Enschede, the Netherlands
anijholt@cs.utwente.nl

ABSTRACT

We discuss our research on designing, implementing and using a virtual (VRML) environment where visitors of the main local theatres in our city can get information about performances and performing artists. Presently visitors can talk to theatre employees (agents) in order to obtain this information and they can get help from a navigation agent in order to explore visualized information (the theatre building, seats that are available, previews of performances). The virtual environment has a multi-agent platform for defining communicating agents and the interaction is multi-modal (keyboard natural language, speech synthesis and recognition, menu items, clickable objects). Current and future research aims at extending this environment to a multi-user environment where users can discuss performances, advise others and can employ multi-media retrieval tools. We shortly survey more fundamental research (done by ourselves and others) that is needed to realize this environment. In particular we look at communication with and between (embodied) agents (representing both users and theatre employees).

Keywords: Virtual Reality, Multi-agents, Multi-modal Interaction, Web Theatre Community.

1 INTRODUCTION

We discuss a virtual world for presenting information and allowing natural interactions about performances, associated artists and groups, availability of tickets, etc., for some existing theatres in the city of Enschede, the Netherlands. The city (having about 150.000 inhabitants) is considered to be the cultural and scientific center of the eastern part of the Netherlands. There are about 15.000 students and there is an interesting co-operation between educational institutes, companies, artist organizations and local and regional governmental organizations.

The main theatre, the so-called ‘Muziekcentrum’, offers its potential visitors information about performances (music, cabaret, theatre, opera) by means of a brochure that is published once a year. In addition to this yearly brochure it is possible to get information at an information desk in the theatre (during office hours), to get (more recent and updated) information by phone (either by talking to a theatre employee or by using Interactive Voice Response Technology) and to get information from local daily and weekly papers and monthly announcements issued by the theatre. The central database of the theatre holds the information that is available at the beginning of the ‘theatre season’. Our aim is to make this information about theatre and performances much more accessible to the general audience.

In our virtual environment the interactions between user (the visitor) and system take place using different task-oriented agents. These agents allow mouse and keyboard input, but interactions can also take place using speech and language input. In the current system both sequential and simultaneous multi-modal input is possible. There is also multi-modal (both sequential and simultaneous) output available. The system presents its information through agents that use tables, chat windows, natural language, speech and a talking face. At this moment this talking face uses speech synthesis with associated lip movements. Other facial animations are possible (movements of head, eyes, eyebrows, eyelids and some changes in face color). These possibilities have been designed and in the design associated with utterances of user or system, but not yet fully implemented.

In this paper it is also discussed how our virtual environment can be considered as an interest community and it is shown what further research and development is required to obtain an environment where visitors can retrieve information about artists, authors and performances, can discuss performances with others and can be provided with information and contacts in accordance with their preferences. In addition, but this has not been realized yet, we would like to offer the environment to the general audience to organize performances, meetings and to present (video) art. The virtual environment we consider is web-based and the interaction modalities that we consider confine to standards that are available or that are being developed for world wide web. For that reason it is necessary to constantly update our environment to anticipate developments on web standards, MPEG, VRML (language and browsers), Java3D, JavaSpeech, etc.

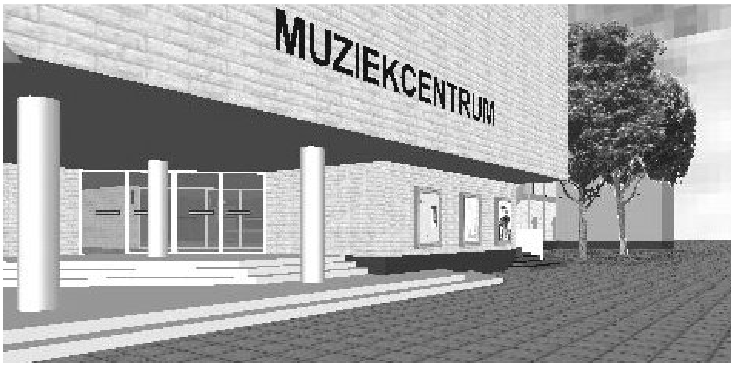

Figure 1: Entrance of the Theatre

From a more global point of view research topics that have been aimed at are:

- Modeling effective multi-modal interactions between humans and computers, with an emphasis on the use of speech and language;
- Commercial transactions, (local, regional and global) information services, collaboration and communication between 'naive' users, education and entertainment in virtual environments.

2 HISTORY AND MOTIVATION

Some years ago, the Parlevink Research Group of the University of Twente started research and development in the area of the processing of (natural language) dialogues between humans and computers. In order to do so, one can choose to take a general, domain-independent approach. This allows general research in syntactic analysis, semantic and pragmatic interpretation and the modeling of dialogues in general. Hence, research has to embedded in current state-of-the-art research on parsing, unification, grammar formalisms, semantics and representation of dialogue utterances, discourse representation and the representation of 'common sense' and world knowledge. That is, knowledge that has to be represented and made accessible in order to get our system to understand user utterances and to generate intelligent system utterances.

Our research led to the development of a (keyboard-driven) natural language accessible information system (SCHISMA), able to inform users about theatre performances and to allow users to make reservations for performances. The system made use of the database of performances in the local theatres of the city of Enschede. In the next sections we will give more information about the design of this theatre information system. The system is far from perfect. However, if a user really wants to get information and has a little patience with the system, he or she is able to get this information. We do not really disagree with a view where users are expected to adapt to a system. On the other hand, wouldn't it be much more attractive (and interesting from a research point of view) to be able to offer environments, preferably on worldwide web, where different users have different assumptions about the available information and transaction possibilities, have different goals when accessing the environment and have different abilities and experiences when accessing and exploring such an environment? We like to offer a system such that we can stimulate and expect users to adapt to it and find effective, efficient, but most of all enjoyable ways to get or to get done what they want, either by themselves, with the help of theatre agents or with the help of others that visit the environment.

In the next section we first discuss how we have added 'context' to our dialogue system. With 'context' we mean that we would like to add visual and auditory cues in the presentation of information and to allow users to choose the (combination of) interaction modalities that best suits his or her preferences for performing the 'task' that has to be done. With 'context' we also refer to the possibility that users come to consider the environment as an interest community, where they can exchange information with other users.

3 VR Embedded Interaction

3.1 Environment Visualization

Our virtual theatre[1] has been built according to the design drawings made by the architects of our local theatre. Part of the building has been realized by converting AutoCAD drawings to VRML97. Video recordings and photographs have been used to add 'textures' to walls, floors, etc. Sensor nodes in the virtual environment activate animations (opening doors) or start events (entering a dialogue mode, playing music, moving spotlights, etc.). Visitors can explore the environment of the building, hear the carillon of a nearby church, look at a neighboring pub and movie theatre, etc. and they can enter the theatre (cf. Figure 1) and walk around, visit the concert hall, admire the paintings on the walls, go to the balconies and, take a seat in order to get a view of the stage from that particular location. When the performance hall is entered, the lights dim, spot lights are moving over the stage and some music starts playing. Information about today's performances is available on an information board (Figure 2) that is automatically updated using information from the database with performances. In addition, as may be expected, visitors may go to the information desk in the theatre, see previews of performances and start a dialogue with an information and transaction agent called 'Karen'. Karen has a 3D animated talking face (Figure 3).

Apart from navigating, clicking on interesting objects (resulting in access to web pages with information about performances, access to web magazines, etc.) and interacting with person-like agents we allow a few other interactions between visitors and virtual objects. For example, using the mouse, the visitor can play with the spotlights and play notes on a keyboard that is standing in some far away part of the building. There is a floor map near the information desk where people can click on positions in order to be 'transported' to their seat in the performance hall so they can see the view they have. On the desk is also a monitor on which they can see pictures or video previews of performances. Unfortunately, most performances do no have a video preview available yet, so we can not display them for every performance that is in the database.

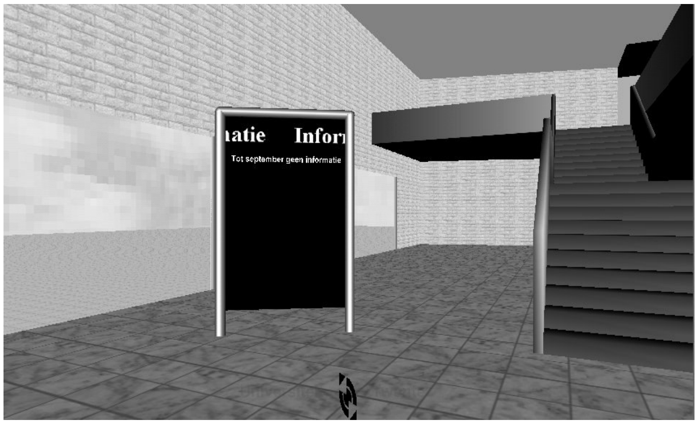

Figure 2: The Information Board

3.2 Visualizing Agents

We assign natural tasks in our environment to agents. It can be useful to visualize them using talking faces and animated 3D avatars. From several studies (cf. Friedman [7]) it has become clear that people engage in social behavior toward machines. It is also well known that users respond differently to different 'computer personalities'. It is possible to influence the user's willingness to continue working even if the system's performance is not perfect. Users can be made to enjoy the interaction and they can be made to perform better, all depending on the way the interface and the interaction strategy have been designed.

In experiments it has been shown that people display different behavior when interacting with a talking face than they do with a text-display interface. This behavior is influenced by the facial appearance and the facial expressions that are shown. People tend to present themselves in a more positive light to a talking face display and they are more attentive when a task is presented by a talking face. From these observations we conclude that introducing talking faces, as we did for Karen and will do for other agents, can help make interaction more natural and shortcomings of technology more acceptable to users.

There is another reason to introduce visualized task-oriented agents. The use of speech technology in

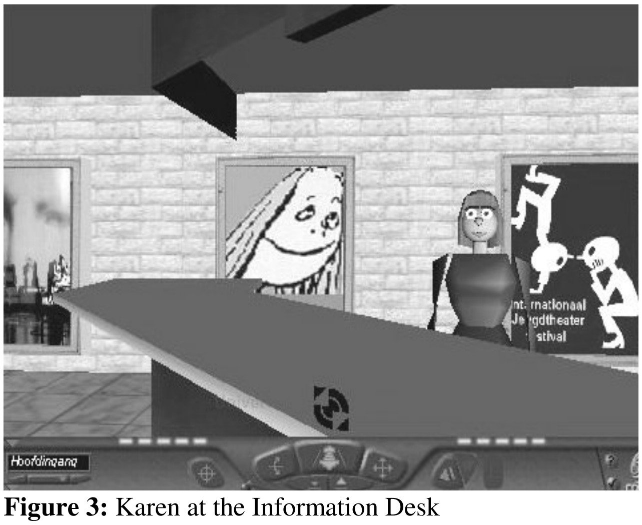

Figure 3: Karen at the Information Desk

information systems will continue to increase. Most currently installed information systems that work with speech, are telephone-based systems where callers can get information by speaking aloud some short commands. Dialogue systems wherein people can say normal phrases become more and more common, but one of the problems in this kind of systems is the limitation of the context. As long as the context is narrow they perform well, but wide contexts are causing problems. One reason to introduce task-oriented agents is to restrict user expectations and utterances to the different tasks for which agents are responsible. Obviously, this can be enhanced if the visualization of the agents helps to recognize the agents tasks.[2]

3.3 MULTI-MODAL ACCESS

When a user has the possibility to change easily from one modality to an other, or can use combinations of modalities when interacting with an information system, then it is also more easy to deal with shortcomings of some particular modality. Multi-modality has two directions. That is, the system should be able to present multi-media information and it should allow the user to use different input modalities in order to communicate with the system. Not all communication devices that are currently available for information access, exploration of information and for transaction allow more than one modality for input or output. This is especially true if we look at world wide web interfaces.

When we look at multi-modal human-computer interaction it is clear that hardly any research has been done to distinguish discourse and dialogue phenomena, let alone to model them, for multi-modal tasks. The same holds for approaches to funnel information conveyed via multiple modalities into and out of a single underlying representation of meaning to be communicated (the cross-media information fusion problem). Similarly, on the output side, there is the information-to-media allocation problem.

Our second observation, certainly not independent from the observation above on modalities for access, exploration and presentation, deals with the actors in a system that has to deal with presenting information, reasoning about information, communicating between actors in the system and realizing transactions (e.g. through negotiation) between actors in the system. For that reason, in addition to a multi-modality approach, there is also a need for a multi-agent approach, where agents can take roles ranging from presenting windows on a screen, reasoning about information that might be interesting for a particular user, and being recognizable (and probably visible) as being able to perform certain tasks.

3.4 VIRTUAL COMMUNITIES

Today there are examples of virtual spaces that are visited and inhabited by people sharing common interests. These spaces can for example, represent offices, shops, class rooms, companies, etc. However, it is also possible to design virtual spaces that are devoted to certain themes and are tuned to users (visitors) interested in that theme or to users (visitors) that not necessarily share common (professional, recreational or educational) interests, but share some common conditions (driving a car, being in hospital for some period, having the same therapy, belonging to the same political party, etc.).

In the previous subsections we have looked at possibilities for theatre visitors to access information, to communicate with agents designed by the provider of the information system and to explore an environment with the goal to find information or to find possibilities to enter into some transaction. Hence, we have a community of people interested in theatre, in music, in performers and their environment has been modeled along the lines of an existing theatre. We need to investigate how we can allow communication between users or visitors of this web-based information and transaction system. Users can help each other to find certain information, they can inform each other (especially when they know about the other's interests), they can have conversations about common interests and they can have domain-related collaboration (e.g., in our case, they can decide to perform a certain play where the actors are distributed among different web sites but sharing the same virtual stage).

As a (not too complicated) example we mention a virtual world developed by the virtual worlds group of Microsoft in co-operation with The Fred Hutchinson Cancer Research Center in Seattle. This so-called "Hutch World" enables people struggling with cancer to obtain information and interact with others facing similar challenges. Patients, families and friends can enter the password protected three-dimensional world (a rendering of the actual outpatient lobby), get

[2] It may be the case that different specialist agents are taken more seriously than one generalist agent. See Nass et al. [11] who report about different appreciation of television programs depending on whether they were presented in a 'specialist' or in a 'generalist' setting.

information at a reception desk, visit a virtual gift shop, etc. Each participant obtains an avatar representation. Participants can engage in public chat discussions or invitation-only meetings. A library can be visited, its resources can be used and participants can enter an auditorium to view presentations.

4 AGENTS IN THE VIRTUAL THEATRE

4.1 AN AGENT PLATFORM IN THE VIRTUAL ENVIRONMENT

In the current prototype version of the virtual theatre we have an information and transaction agent, we have a navigation agent and there are some agents under development. An agent platform has been developed in JAVA to allow the definition and creation of intelligent agents. Users can communicate with agents using speech and natural language keyboard input. Any agent can start up other agents and receive and carry out orders of other agents. Questions of users can be communicated to other agents and agents can be informed about each other's internal state. Both the information & transaction agent and the navigation agent are in the platform. But also the information board, presenting today's performances, has been introduced as an agent. And so can other objects in the environment.

It is an important question how to integrate the human visitors of our environment with our models of agent interaction, with our models of multi-modal interaction and multi-media presentation, with models of non-verbal agent behavior (associated with verbal behavior) and with models of agent movements. We will return to this question in forthcoming sections, but it should be mentioned that hardly any research results are available and that no experiments have been performed from which we can learn how humans behave in such agent-rich environments.

4.2 THE INFORMATION & TRANSACTION AGENT

Karen, the information & transaction agent, allows a natural language dialogue with the system about performances, artists, dates, prices, etc. Karen wants to give information and to sell tickets. Karen is fed from a database that contains all the information about performances in the (existing) theatre.

Our current version of the dialogue system of which Karen is the face is called THIS v1.0 (Theatre Information System). The approach used can be summarized as 'rewrite and understand' (Lie et al. [10]). User utterances are simplified using a great number of rewrite rules. The resulting simple sentences are parsed. The output can be interpreted as a request of a certain type. System response actions are coded as procedures that need certain arguments. Missing arguments are subsequently asked for. The system is modular, where each 'module' corresponds to a topic in the task domain. For example, a module has to take care of a date a user is referring to (next Wednesday, over two weeks, tomorrow).

Presently the input to Karen is keyboard-driven natural language and the output in our for the general audience WWW accessible virtual world is screen and menu based. In a prototype system we allow Karen to use a mix of speech synthesis and information presentation on the screen. As mentioned earlier, in this prototype system Karen's spoken dialogue contribution is presented by visual speech, that is, a 'talking face' on the screen, embedded in the virtual world, mouths the questions and part of the responses. If necessary, information is given in a window on the screen, e.g., a list of performances or a review of a particular performance. The user can click on items to get more information or can type in further questions concerning the items that are shown.

4.3 THE NAVIGATION AGENT

Navigation in virtual worlds is a well known problem. Usually, navigation input is done with keyboard and mouse. This input allows the user to move and to rotate, to jump from one location to an other, to interact with objects and to trigger them. We developed a navigation agent that helps the user to explore the environment and to interact with objects by means of speech commands. The navigation agent knows about the user's coordinates in the virtual world and it has knowledge of the coordinates of a number of objects and locations. This knowledge is necessary when a visitor refers to an object close to the navigation agent in order to have a starting point for a walk in the theatre and when the visitor specifies an object or location as the goal of a route in the environment. The navigation agent is able to determine its position with respect to nearby objects and locations and can compute a walk from this position to a position with coordinates close to the goal of the walk.

Verbal navigation requires that names have to be associated with different parts of the building, objects and agents. Users may use different words to designate them, including references that have to be resolved in a reasoning process. The current agent is able to understand command-like speech or keyboard input. It hardly knows how to communicate with a visitor. The phrases to be recognized must contain an action (go to, tell me) and a target (information desk, synthesizer). It tries to recognize the name of a location in the visitor's utterance. When the recognition is successful, the agent guides the visitor to this location. When the visitor's utterance is about performances the navigation agent makes an attempt to contact Karen, the information and transaction agent. In progress is an implementation of the navigation agent in which it knows about (or should be able to compute) what is in the eyesight of the visitor, focus of gaze, some history of visits and interactions, etc.

We haven't yet decided whether to make our navigation agent visible as an avatar. How will the visitor behave? Consider the avatar as a representation of him or her self? Consider the avatar as an agent provided by the environment? Obviously, this depends on the way we visualize this agent. Does it have our own face and body or is it Lara Croft that walks in front of us?

4.4 LANGUAGE SKILLS OF THE AGENTS

At this moment our agents have different language skills. On the one hand we have Karen and a grammar specification of the input for Karen based on a corpus of WoZ obtained keyboard-based dialogue utterances. On the other hand we have a navigation agent with language skills that are based on the current limitations of speech input. We would like to automate the process of assigning language skills to agents in our environment as much as possible. Therefore we hope to see speech recognition technology move forward from keyword spotting, to finite state utterance specification to (word-graph) based context-free language specification. At this moment we follow the developments of Philips speech recognition software when looking at the recognition of spoken dialogue utterances for different agents. More fundamental, however, is our approach to induce grammars (context-free, probabilistic, unification constraints) from corpora of utterances (see Ter Doest [6]) collected in Wizard of Oz experiments. In this approach we tag a corpus with syntactic categories and superficial structure using Standard Generalized Markup Language (SGML). From this tagged data grammar rules, unification constraints and probabilities are derived. We have tested grammars on 'seen' and 'unseen' data from Karen's same domain using a probabilistic left-corner parser for PATR II unification grammars. In a similar way we have induced for our navigation agent a probabilistic grammar from a corpus of user utterances that have been obtained from several scenarios presented to (potential) visitors from the theatre. This grammar is a start. It allows the design of a primitive system and it allows bootstrapping this system from the original corpus and from corpora obtained from logging the interactions between visitors and the navigation agent. Clearly, this approach still requires the integration of speech recognition technology with natural language specification and understanding. For that reason it may be useful to investigate the generation of finite state probabilistic (unification) grammars from corpora of utterances.

In our current web-based system we have speech recognition on the server side. This requires the recording of commands on the client side and a robust transporting of the audio files. It does not require users to install speech recognition software or to download a speech recognition module as part of the virtual world from the server. Users do need however audio-software, which is usually available anyway. For speech recognition we use Speech Pearl, commercial speech understanding software from Philips. Recognition is based on keyword spotting. A next version of the software will allow a finite state specification of the user's input for speech recognition.

For text-to-speech synthesis we use the Fluent Dutch Text-to-Speech system (Dirksen [5]). It runs on top of the MBROLA diphone synthesizer. It uses a male Dutch voice, a female version of the system is in development. Lexical resources of *Van Dale Dictionaries* have been used to obtain phonetic descriptions of the words.

4.5 AGENTS THAT SELL, ADVISE, BUY, . . .

When we look at Karen, it will be clear that her boss would like her to sell as many tickets as possible. A theatre director will have certain preferences in choosing performances for her theatre, but when they have been chosen the aim is to have the performances sold out every night. How can we program Karen (and maybe also the navigation agent) such that the behavior towards a visitor increases the chance to reach this 'private' goal?

More generally, we can imagine that web pages and virtual environments will become inhabited by sales agents, hosts, guides, etcetera that offer information and answer questions (and try to sell products) in a way similar to Karen or our navigation agent, but also in a way that allows building up relationships with users through social "chit-chat" about family, work or sports.

That is, these agents should have specific knowledge about a certain domain, a domain which may range from detailed knowledge about Shakespeare, performances in a next season and cars in a showroom to drinks that are available in a bar. But, in addition to that, they have some superficial global knowledge that allows them to give socially appropriate answers to keep up a conversation about topics they hardly know of and they have some special hobbyhorses and some specific knowledge with which they can steer a conversation and which makes them 'believable' to the user.

One of our concerns in the near future will be the introduction of such properties in our present agents. Maybe Karen shouldn't be allowed to give too personal opinions about performances and artists, but some more human-like conversational behavior should be considered. Moreover, if a user really wants to know or to exchange opinions about performances and artists she should be able to communicate with other visitors of the environment or be able to address agents employed by the theatre who can share their specific knowledge (embedded in some kind of social conversation) with the visitor.

Agents that allow human-like conversation have been designed, both in research and in commercial environments. An example of a virtual web agent is Jennifer James, designed by *Extempo Systems Inc*. She sells cars in a virtual auto show room, 24 hours a day. Jennifer has a history in racing, which is useful in talking with customers. She wants to know about a customer's background and preferences. Questions are asked in a friendly and sometimes ironic way. Jennifer knows about cars and racing, but questions or comments on family or country music can be dealt with in a believable way. Jennifer has been visualized as a smiling saleswomen dressed in a red and white jump suite. Jennifer has personality, customers feel confident to talk with her, they give information about themselves and a company that employs Jennifer will build relationships of affection, trust and loyalty with its customers. The information that is elicitated from a customer is stored in marketing/customer databases.

Jennifer makes a charming attempt to understand natural language and to interact with objects in the scene. The customer uses the keyboard to communicate with Jennifer. Her body and face are real-time animated, where animations attempt to be consistent with role and personality, and with the events of the dialogue. Jennifer talks back using speech synthesis.

More examples of agents that display human and social qualities have been introduced. Of course, we can go back to 'conversational agents' such as Weizenbaum's Eliza and Colby's Parry (both from the sixties) or Julia, a chat box character of the early nineties. However, these characters have not been given the task to inform visitors, to give information, to guide and help or to purposely try to seduce potential buyers to buy commercial products, to visit sites and events and to take part in activities.

5 Visual Speech, Facial Animation, Gestures and Movements

5.1 Introduction

Our agent platform allows the introduction of new agents. The interaction that is allowed between agents is primitive, but it nevertheless allows to have a change of control from navigation agent to information agent and vice versa. The agents don't have an explicit BDI model, rather their beliefs, desires and intentions are hidden in their dialogue intelligence. This needs to be changed in future implementations in order to be able to maintain the environment when other agents will be introduced and when users themselves get the opportunity to introduce agents (for example, themselves). For the agents offered by the environment we require that they have a certain intelligence and that they can display some verbal and non-verbal behavior. They can also address each other, in order to satisfy certain wishes of the visitors or of the creators (owners) of the environment.

We may have situations where both agents in an dialogue represent human participants, where one of the participants is human and the other is synthetic, and where both are synthetic. Obviously, rather than have a dialogue between two agents, we can have interactions involving three or more human and synthetic participants. In a shared environment some agents can decide or can be asked to help an other agent or to collaborate in order to perform a certain task. The results of the collaboration can become observable (visible, audible, ...) for themselves, for one or several other agents (not necessarily involved in the collaboration) or for the general audience that visits the virtual environment. In our environment this will amount to noticing that some activity is taking place (e.g., agents get together to have a jam session), that the history of the environment has been changed (a jam session has been added to the history), that the environment itself has been changed (instruments have been moved from one place to an other) or that the state or knowledge of some agents have been changed (they have learned preferences of other players and how to deal with these preferences during a joint performance).

Clearly, it is much too ambitious to make an attempt to implement an environment in which we allow all such activities. At this moment, in our 'laboratory' environment, we concentrate on research on modeling verbal and nonverbal behavior of agents (in particular behavior that shows in their faces) with the aim to obtain research results that can be used to model interactions between agents, between agents and users, and between users, in commercial, educational and cultural interaction.

5.2 FACING THE INFORMATION AGENT

We developed some virtual faces in a 3D-design environment (Berk [2]). 3D data can be converted to VRML-data that can be used for viewing and animation of a virtual face. A picture of a real human face can be mapped onto a virtual face. We are researching various kinds of faces to determine which can be best used for our applications. Some are rather realistic and some are more in a cartoon-style (cf. Figure 4). The information agent has been given a virtual 3D face. The face is capable of visualizing the speech synchronously to the speech output. This involves lip-movements according to a couple of visemes. We also have defined facial expressions according to user's input or system's output.

A dialogue window is shown when users approach the information-desk while they are navigating in the virtual theatre. This window, the JAVA Schisma applet, is available to formulate questions or to give answers to the system's questions. The user types the questions on a keyboard in Dutch sentences. The answers to the questions are to be determined on the server side: the Schisma server. Answers or clarifying questions are passed to the JAVA Visual Speech Server Application on the server side. This application filters the textual output of the dialogue system in parts that are to be shown in a table or a dialogue window and parts that have to be converted to speech. The parts that are to be shown in the dialogue window or a table, like lengthy descriptions of particular shows or lists of plays are send to the Schisma Client Applet where they are shown on the screen. The parts of the Schisma output that are to be spoken by the virtual face are converted to speech with the Text-to-Speech Server. The input is the raw text and the output is the audio file of this spoken text and information about the phonemes in the text and their duration.

Figure 4 A Cartoon Face

How do we control the responses, the prosody and the artificial face? Response actions are combinations of basic domain related actions (e.g. database queries) and dialogue acts to convey the results of the query. Dialogue acts describe the intended meaning of an utterance or gesture. The response generation module selects a way to express it. It determines utterance-structure, wording, and prosody of the system response. In addition it controls the orientation and expression of the face, the eyes, and the coordination of sounds and lip movement. For details of the design of the response module see Nijholt and Hulstijn [12]. For the approach we use for gaze behavior we refer to Cassell et al. [4].

Presently, we are doing experiments with an eye tracker system, where knowing where the visitor is looking at is detected by an infrared camera. On top of this camera is an infrared source projecting invisible light into the eye. This light is reflected by the retina and the cornea of the eye. These reflections make it possible to determine where a person is looking at. In particular, it is possible to determine to which avatar a person is looking. This allows management of multi-user conversations in a virtual environment, where each user knows when and which other users are looking at him or her. This leaves to a certain degree open how the user is represented in the environment, but at least user gaze directional information can be conveyed. This approach allows visitors of our environment to address different task-oriented agents in such a way that speech recognition and language understanding are tuned to the particular task of the agent; therefore quality of recognition and understanding can increase considerably, since the agent may assume that words come from a particular domain and that language use is more or less restricted to this domain. That is, we can restrict lexicon and language model to the utterances that are reasonable given the agent. Obviously, we should try to visualize agents in such a way that it is clear from their appearance what they're responsible for and what a visitor can ask them. An attempt should be made to ensure that any agent is able to determine that she isn't the right agent to answer a visitor's questions and therefore should direct the visitor to an other task-oriented agent or to an agent having global knowledge of the task-oriented knowledge of the other agents in the virtual environment.

5.3 Naturally Moving Animated Agents

If we want agents that are visible for the visitor, agents that walk, agents that show how to do certain things, agents that perform, then we need rather natural visualization of movements of agents (movements of body, legs, arms, fingers, etc.) and animation of facial expressions, all in accordance with the tasks that the agent has to perform and the interaction with the visitor (if that is required). We have the following process in mind for the creation of naturally moving animated agents. First, it would be the modeling of agents, then the modeling of movements, then the control system. At each step we must take into account the goals of the next step(s) so the different steps can be used together. For example: the agents should be modeled in mind with the fact that they must be animated later, and the animation sequences must be directed by a control system. In the case of modeling and animating the best approach will be to build entire systems to have the capability to build and experiment with multiple agents and movement types.

The main difficulty with animated agents is the animation itself. An agent can be modeled with quite a few possibilities, as can be observed from widespread modeling programs and packages. The primary task should be therefore to identify from these modalities the ones applicable in the environment. They need to respond properly to the deformations of the agent's body such as bending, twisting of the joints, taking different pose, moving body parts. All this in order to ensure that the body of the agent and the movements of the agent appear natural, free from any distortion or lack of continuity. This alternative should form the technical background for visualizing the agent(s) together with the procedural and scriptable properties of a virtual environment. Such properties as scripting and procedures are welcome and necessary for defining the different phases of the movement, and at the same time, keeping the animation data at minimal size and the frame rate at high values.

Movements can be assembled from movement primitives These primitives are not so many in numbers and once defined they can be combined to generate a wide range of behavior. Even if all the primitives have to be defined, the movement data they need to contain will be still short and usable through WWW. The data should be given in a key-frame format as opposed to full motion path specification which needs special equipment for capture and also needs more bandwidth or opposed to goal-directed, constraint based and algorithms controlled movement systems that need high computational power. The key-frame approach is somewhere in the middle, taking just a minimal set of motion data and the gaps between the motion data can be filled using interpolation by the rendering program (browser) itself, in function of the rendering speed it can achieve. The animation description/data due to its shortness can be stored together with the model, and therefore stored locally after downloading. This way the environment has the possibility to generate the movements from the local data quicker, not having to wait for the data to download.

Our first task is to determine the best modeling possibility, in concordance with the motion animation possibilities. Second, to build an editor based on the agent model used. Third, to build a motion editor to define the primitives and to combine them into 'actions'. Fourth, to build the control system directing the agent, which will be the 'brain' of the agent.

6 MM-Retrieval in the VR Environment

Presently we investigate how to store, index and retrieve theatre-related multi-media objects (text, pictures, audio, video) in such a way that a visitor of our virtual theatre environment can address Karen in such a way that she is not only able to inform the user about artists, performances, dates and available seats, but also knows how to provide the user with pictures, audio and video fragments (previews) of performances. In general, if a visitor asks about a particular performance, it is not difficult to let Karen present the (available) associated multi-media information. However, if the visitor's questions are about or address associated information rather than explicitly making references to a particular performance, then our theatre agents have to start reasoning with the information that is available in the database and information that is available in a knowledge representation scheme in which general knowledge about the theatre domain has been stored.

At this moment no explicit, comprehensive and suitable ontology of the theatre information and booking domain has been designed. Karen knows about this domain, but her knowledge is hidden in the linguistic (syntactic, semantic) and the dialogue management (pragmatic) knowledge of the present system. This knowledge nevertheless allows a better mapping on the information that is available in the theatre database. However, other agents have no access to this implicit knowledge. Similarly, our navigation agent knows about the geography of the theatre and how users generally ask questions about the theatre. However, knowledge about the geography has been represented as a list of coordinates of particular locations and possible ways to refer to them. Hence, the user has some freedom how to address these locations. However, whatever synonyms are allowed, the list is predefined and only the navigation agent knows how to access this information. Our speech recognition software, that is, the number of phrases it can recognize reasonably correctly in normal circumstances, does very much determine the navigation's agent intelligence.

How to handle a user's question about artists and performances, where the user doesn't use names that are available in the theatre database? That is, can we provide Karen with general knowledge about artists and performances which goes beyond the information that is available in the database? , and, if it turns out that the user doesn't know about Karen's knowledge and such knowledge is not complete, can we nevertheless map the user's question to a database query?

As an example, suppose someone asks: "Well, what's her name, Michelle, eh, I know she received an Oscar last week. Are there any movies with her showing this week?" At this moment we think that it is reasonable to expect from our system that it starts searching WWW with a question that has as keywords: 'Oscar' and 'Michelle'. The intelligent theatre search engine should deliver the name of the actress and this should be sufficient to search the database for movies showing this week (or maybe later this month) with this particular actress in one of the roles.

A knowledge representation network with the main concepts and their relationships in the world of theatre (opera, musicals, plays, music, cabaret, etc.) and theatre actors (directors, performers, technicians, reviewers, etc.) has to be designed in order to connect information available on WWW and information available in local databases through a process of reasoning.

Textual information is important, since it is widely available. Newspapers, magazines and WWW pages contain reviews about performances, movies and performers. However, there is no need to confine ourselves to textual information. Presently, our research group is involved in some national and European projects on indexing and retrieval of multi-media information, including text, lay-out, pictures and captions, video subtitles, transcripts of spoken movie dialogues and spoken language. Integrating this research on multimedia indexing and retrieval and natural language (spoken) dialogue systems embedded in environments where users can 'say what they see' (ranging from handheld devices to immersive virtual reality environments) is a topic high on our research agenda. Since we think it is fruitful to consider WWW as an extension of our theatre databases we need an intelligent domain agent that knows how to filter and present the results of a WWW information search to a naive user looking for some information that may help in having a nice evening.

7 SOFTWARE ENGINEERING ISSUES

7.1 USE CASES

In our system we have to deal with different agents and with different dialogues or interactions between agents and users (visitors of our theatre). We are investigating the use of object-oriented and agent-oriented design techniques to build our theatre. One design approach which we found useful is the description of use cases. A use case is a detailed account of the context and goals of the different actors involved in a transaction, together with a possible sequence of actions. They are recommended in object-oriented development to find common data elements to be modeled as objects. Use cases become part of the requirement specification document that serves as a contract between the system developer and the customer, in this case the service provider. Because they can be intuitively understood, use cases facilitate communication with user focus groups and with the customer. Based on several scenarios for the navigation and the information and transaction agent use cases have been developed (Hulstijn [9]).

7.2 FORMAL MODELING OF INTERACTIONS IN VIRTUAL ENVIRONMENTS

Both from an ergonomical and a software-engineering viewpoint, the design of interaction in virtual environments is complex. 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 new line of research we have taken is an attempt to address both of these issues by means of a formal modeling technique that is based on the process algebra

CSP (Communicating Sequential Processes). For that reason, in our virtual theatre 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 (such as video versus audio and text versus graphics) may also be modeled as separate channels.

This modeling technique has some strong points. Firstly, and most generally, such a simplified and formal model enables a clear and unambiguous specification of system architecture and dynamics. Secondly, 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 explicitly showing which options are available when and through which modalities. Thirdly, it enables automatic prototyping, such as architecture visualization and verification of some system properties.

For more details about this approach we refer to Schooten et al. [14]. There it is also shown how a CSP description can be coupled to a simplified user interface and executed, so that the specified system can be tried out immediately. Specifications map closely to software architecture, reducing the cost of building a full prototype.

8 VIRTUAL PERFORMANCES

Now that we have a virtual theatre where people can look around and get information on performances, wouldn't it be nice to apply this virtual reality environment to other theater-related purposes? There exist research projects aiming at providing professionals tools and environments to help in pre-producing performances. That is, allowing users to build a scenography of a performance, to move through virtual models of stage sets in real time, to experiment with lights or camera effects, change points of view, etc. In Rodriguez et al. [13] a virtual stage is offered to choreographers where they can preview a performance using animated human figures.

It is not unusual today to have meetings in virtual environments. Lectures have been organized in chat environments and meetings have been held in visualized meeting places. Online performances have been given, including a *Hamlet* parody on IRC and *The Odyssey* by Homer. An other example is Shakespeare's *A Midsummer Night's Dream*, a VRML production performed live on April 26, 1998 (see [15]). A virtual concert was held on July 4th of 1999, with live musicians and streaming audio, occurring on WWW. The musicians at this event were represented as avatars. A musician could play his own music and talk about it with the audience. However, musicians just played their own music and no possibility was offered to have musicians at different websites playing together.

At the Università degli Studi di Milano a 3D model of the Teatro alla Scala in Milano has been constructed [1]. In addition research is done on baroque dance animation with virtual dancers [3]. 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 dancers in our theatre.

9 CONCLUSIONS AND FUTURE RESEARCH

In this paper we reported about on-going research and it is clear that all issues that have been discussed here need further research. We intend to continue with the interaction between experimenting with the virtual environment (adding agents and interaction modalities) and theoretic research on multi-modality, formal modeling, natural language and dialogue management. In particular the integration of visitors in an agent platform that models a uniform verbal and non-verbal behavior is required in order to be able to maintain and extend the virtual environment.

As may have become clear from the previous sections, our approach to designing a virtual environment for an interest community is bottom-up. At this moment the system has two embodied agents with different tasks and with no interactions between them. Moreover, the agents do not employ a model of a user or of user groups. In general, when we talk about interface agents we mean software agents with a user model, that is, a user model programmed in the agent by the user, provided as a knowledge base by a knowledge engineer or obtained and maintained by a learning procedure from the user and customized according to his preferences and habits and to the history of interaction with the system. In this way we have agents that make personalized suggestions (e.g. about articles, performances, etc.) based on social filtering (look at others who seem to have similar preferences) or content filtering (detect patterns, e.g. keywords) of the items that turn out to be of interest to the visitor. One of our aims is to provide visitors with personal assistants ('butlers') that know about the visitors' preferences, that can exchange information with other personal assistants and that can search for and filter information that is of interest for the visitor. In particular we hope to integrate research that comes available from the European projects *Magic Lounge* (which aims at developing tools that allow intelligent communication services in virtual meeting places; these tools include shared white boards and chat environments, but also tools that record and store interaction events such that it becomes possible to browse earlier interactions and inspect individual contributions) and *PERSONA* (a project about navigation in two- and three-dimensional interfaces; in this project the concept of social navigation is explored, that is, navigation that exploits the possibility to talk to other users and to agents for obtaining information, including the following of their trails in the information space).

In 1999 some versions of our virtual environment have become available for other research groups to work on. For example, in a joint project with the TNO Human Factors Research Institute user evaluation studies will be done and we hope this will help in future decisions about the direction of our work on the theatre information and transaction service interactions and the environment where they take place. A simplified and localized version of the virtual environment has been placed at a Dutch technology activity center (Da Vinci). Here, visitors are allowed to play with the system and their (verbal) interactions with the system are logged for analysis. At WWW our system can be visited at URL http://parlevink.cs.utwente.nl/.

Acknowledgements: The research described here is a collaborative effort of the Parlevink Research Group. We particularly want to mention the work of Eduard and Richard Selles, Hendri Hondorp and Joris Hulstijn.

REFERENCES

[1] Alberti, M.A. & P. Trapani. On the Opera theatre simulation. Proc. *Eurographics '99*: Short papers and demonstrations. September 1999, 114-116.

[2] Berk, M. van den. Visuele spraaksynthese. Master's thesis, University of Twente, 1998.

[3] M. Bertolo, P. Maninett & D. Marini. Baroque dance animation with virtual dancers. Proc. *Eurographics '99*: Short papers and demonstrations. September 1999, 117-120.

[4] Cassell, J. & K.R. Thórisson. The power of a nod and a glance: envelope vs. Emotional feedback in animated conversational agents. *Applied Artificial Intelligence*, to appear.

[5] Dirksen, A. and Menert, L. Fluent Dutch text-to-speech. Technical manual, Fluency Speech Technology/OTS Utrecht, 1997.

[6] Doest, H. ter. *Towards Probabilistic Unification-Based Parsing.* Ph.D. Thesis, University Twente, February 1999.

[7] Friedman, B. (ed.). *Human Values and the Design of Computer Technology*. CSLI Publications, Cambridge University Press, 1997.

[8] Hulstijn, J. & A. van Hessen. Utterance Generation for Transaction Dialogues. Proceedings 5th *International Conf. Spoken Language Processing*, Vol. 4, Sydney, Australia, 1998, 1143-1146.

[9] Hulstijn, J. Modeling usability: Development methods for dialogue systems. In: Alexandersson, J. (ed.) IJCAI-99 Workshop on *Knowledge and Reasoning in Practical Dialogue Systems*, Sweden, 49-56.

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

[11] Nass, C., B. Reeves & G. Leshner. Technology and roles: A tale of two TVs. *Journal of Communication* 46 (2), 121-128.

[12] Nijholt, A. & J. Hulstijn. Multimodal Interactions with Agents in Virtual Worlds. In *Future Directions for Intelligent Systems and Information Science*, N. Kasabov (ed.), Physica-Verlag: Studies in Fuziness and Soft Computing, 1999, to appear.

[13] Rodriguez, I., J. de Pedro & D. Meziat. Integrating virtual humans and virtual stages. In: Proceedings Third World Multiconference on *Systemics, Cybernetics and Informatics (SCI'99)*, Vol.8: Concepts and Applications of Systemics, Cybernetics and Informatics. M. Torres, B. Sanchez & E. Wills (eds.), USA, 1999, 396-401.

[14] Schooten, B. van, O. Donk & J. Zwiers. Modeling interaction in virtual environments using process algebra. In: Proceedings *Interactions in Virtual Worlds (IVW'99).* Twente Workshop on Language Technology 15, University of Twente, May 1999, 195-212.

[15] http://www.vrmldream.com/