Commerce & Entertainment in the Twente Virtual Theatre Environment

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

ABSTRACT

In this paper we discuss research on a virtual theatre environment. The theatre has been built using VRML and therefore it can be accessed through World Wide Web. In the environment we employ several agents. The theatre allows navigation input through keyboard function keys and mouse, but there is also a navigation agent which tries to understand keyboard natural language input and spoken commands. Feedback of the system is given using speech synthesis. We also have Karen, an information agent which allows a natural language dialogue with the user with the help of, among others, text-to-speech synthesis and a talking face. We discuss how this particular environment can be seen as an environment for commerce and entertainment and we discuss further requirements that should be fulfilled by our agents in order to obtain a 'believable' environment.

Keywords: Virtual Reality, Multi-modal Interaction, Agent Technology

1 INTRODUCTION

We present research on the development of a virtual environment in which users can display different behaviors and have goals that emerge during the interaction with this environment. Users who, for example, decide they want to spend an evening outside their home and, while having certain preferences, cannot say in advance where exactly they want to go, whether they first want to have a dinner, whether they want to go to a movie, theatre, or to opera, when they want to go, etc. During the interaction, both goals, possibilities and the way they influence each other become clear. One way to support such users is to give them different interaction modalities and access to multimedia information. We discuss a virtual world for presenting information and allowing natural interactions about performances, artists, availability of tickets, etc., for some existing theatres in our hometown.

The interactions between user (the visitor) and system (a virtual representation of an existing theatre) take place using task-oriented agents. They allow mouse and keyboard input, but interactions can also take place using speech and language input. In the 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). At this moment these possibilities have been designed and in the design associated with utterances of user or system, but not yet fully implemented.

It is also discussed how our virtual environment can be considered as an environment for commerce and entertainment. At this moment information is provided about performances and the products that are sold are performances. There are differences between selling cars and performances, but there are many common characteristics with which we should be able to deal with before going into looking at 'subtle' differences. For example, there is an information phase and a transaction phase in a dialogue, the partners may return to the information phase if the transaction phase is not successful, both buyer and seller may have interests that do not necessarily match, etc. In short, selling tickets for theatre performances is a commercial transaction and we can try to design environment and interactions in ways that aim at increasing the number of tickets sold, at increasing the satisfaction of users or at compromising between the both. Clearly, the environment offers services. When these services are closely related to the selling of tickets one may argue that the costs of using these services should be included in the price of the tickets. However, other services may be offered, e.g. information about performances in other theatres, information about artists in general, and, much more interesting, the possibility to 'buy' a performance for viewing at home. That is, rather than going to the theatre, a customer can decide to stay home and watch a performance (comparably to pay-TV) sitting in his arm-chair at home.

We can also look at our environment from an entertainment point of view. First of all, visiting the environment with the goal to select a certain performance and to get more information about this performance can be fun. It can be compared with 'traveltainment', comparing different holiday travels, choosing a particular travel, looking at pictures or video presentations of the destination, reading about the history of places to be visited, etc. Moreover, traveltainment may also imply visiting evenings with fellow travelers, share information and experiences with others, etc. In our case we may strive for an environment where visitors can retrieve (multimedia) information about artists, authors and performances, can discuss performances with others and can be provided with information and contacts in accordance with their preferences. Secondly, we would like to use our environment (and to offer our environment) to organize performances, meetings and to present art. There have been examples of performances on the web. A visitor can passively consume what is offered (paying and watching), but WWW also offers the possibility that s/he can get involved in the entertainment that is being offered. Audience participation can range from showing appreciation, not only at the end, but also during the performance (e.g., with the aim to influence the story action) to active participation.

Here we introduce the environment which has been designed and implemented for experimenting with commercial and cultural services and with the possibility for a visitor to consume these services and to attribute to them. In the next section (section 2) we will give more details about the virtual environment we built and we discuss the agents that are available to sell tickets and to help the visitor to find its way in the theatre environments. In section 3 we present our research to improve verbal and nonverbal behavior of these agents in their communication with the visitors of our web-based environment. Section 4 of this paper is devoted to the possibility to have visitors and agents performing, contributing to and attending real-time web-performances. In section 5 we present conclusions and future research.

The environment we consider is web-based and the interaction modalities that we consider confine to standards that are available or are being developed for world wide web. For that reason it is necessary to constantly update our environment to anticipate developments on web standards. The topics that we deal with have to do with information and transaction services. It will become clear how to generalize our approaches and how to tune them to domains different from our theatre domain.

2 THE TWENTE VIRTUAL THEATRE ENVIRONMENT

2.1 Visualization of the Environment

Some years ago we started research and development in the area of the processing of (natural language) dialogues between humans and computers. This 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. The system is rather primitive. 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 decided to visualize the environment in which people can get information about theatre performances. Visualization allows users to refer to a visible context and it allows the system to disambiguate user's utterances by making use of this context. Moreover, it allows the system to influence the interaction behavior of the user in such a way that more efficient and natural dialogues with the system become possible.

Figure 1 Entrance of the Virtual Theatre

Our virtual theatre 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 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 music starts playing. Information about today’s performances is available on an information board 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/transaction agent called ‘Karen’, making use of a 3D talking face.

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 isolated 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 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.

The first version of the virtual theatre took about 6.4 Mb. From this 628 Kb was VRML code. Much effort has been taken to optimize the system, reducing it to 5.4 Mb with 375 Kb code. The rest of the reduction was obtained by manipulation of the textures. Things that were done for size reduction include:

- prototyping objects with the same structure and eliminating not visible and therefore unnecessary surfaces
- simplifying the most used or complicated structures (in this case the chairs, the keyboard)
- defining and reusing materials, objects, coordinates
- using JPEG textures instead of the GIF format (except for transparency considerations) and decreasing the color depth of textures that use few colors

2.2 The Information & Transaction Agent

In the current prototype version of the virtual theatre we distinguish between different agents: 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.

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. Developing skills for Karen, in this particular environment, is one of the aims of our research project. This research fits in a context of much more general 'intelligent' (web-based) information and transaction services.

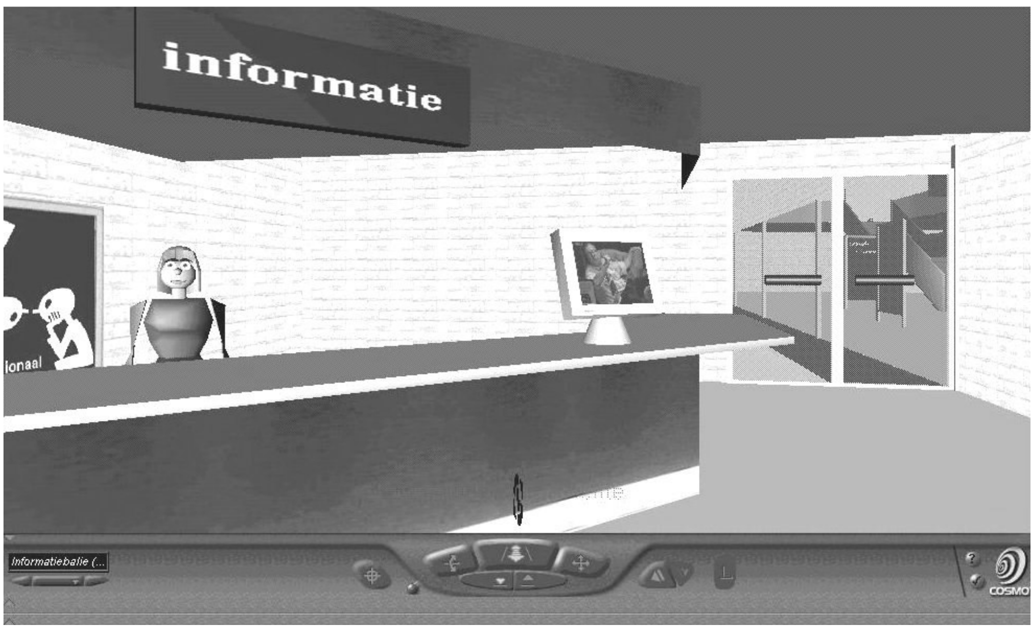

Figure 2 Karen at the Information Desk

The approach used in our current version of the dialogue system can be summarized as 'rewrite and understand'. 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.

There are also modules for each step in the understanding process: the rewriter, the recognizer and the dialogue manager. The rewrite step can be broken down into a number of consecutive steps that each deal with particular types of information, such as names, dates and titles. The dialogue manager initiates the first system utterance and goes on to call the rewriter and recognizer process on the user's response. Also, it provides an interface with the database management system (DBMS). Queries to the database are represented using a standard query language like SQL. Results of queries are represented as bindings to variables, which are stored in the global data-structure, called context. Based on the user utterance, the context and the database, the system has to decide on a response action, consisting of database manipulation and dialogue acts. The arguments for the action are dug out by the dedicated parser. All arguments that are not to be found in the utterance are asked for explicitly.

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 in the introduction, 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.

2.3 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 its own 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 short walk from this position to a position with coordinates close to the goal of the walk.

Navigation also requires that names have to be associated with the different parts of the building, the objects and the 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, keyboard). 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):

- Current position, focus of gaze of the user and what is in the eyesight of the visitor;
- Objects and the properties they have, geometric relations between objects and locations;
- Possible walks towards objects and locations and some knowledge of previously visited locations or routes;
- The action it is performing (or has performed)
- Some knowledge of the previous communication with the visitor.

It is well known that there are conditions that invite users to respond to a computer or to a computer-mediated entity as being human actors in a 'real' world. Is it necessary to distinguish situations and (probably) associated utterances where a visitor addresses the navigation agent as

- a help agent knowing or sharing the visitor's coordinates and camera view;
- an agent which looks at environment and visitor 'from above';
- a personification of him or her self?

In addition, when we decide to make our navigation agent visible as an avatar, how will the visitor behave? Consider the avatar anyway as a full representation of him or her self? Consider the avatar as an agent provided by the environment (more or less tuned to a possible implicit identification)? Obviously, this depends on the way we visualize this agent. Does it have our own face and body or is it Lara Croft walking in front of us? In the latter case it is not at all strange that a visitor, who has a different camera view, issues a warning, saying: "Look behind you." In the case that the avatar is considered to be a personification of the visitor this is a rather schizophrenic utterance. Our current agent is not visible and so it only resembles Lara Croft in a visitor's imagination.

2.4 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. It uses 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. If we continue to use this software we should decide about the possibility to model user utterances in navigation dialogues on finite state grammars or to distinguish the navigation language model for speech recognition from the navigation language model for speech utterance understanding. In the latter case it is preferable to have a word graph recognition from the speech recognition module of the system.

More interesting, we think, is our approach to induce grammars (context-free, probabilistic, unification constraints) from a corpus (see [1]). Presently 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 of 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.

In the design of utterance generation by the information agent a list of utterance templates is used. Templates contain gaps to be filled with information items: attribute-value pairs labeled with syntactic and lexical features. Templates are selected on the basis of five parameters: utterance type, the body of the template and possible empty lists of information items that are to be marked as given, wanted and new. The utterance type and body determine the word-order and the main intonation contour. The presence and number of information items in the given, wanted and new slots, as well as special features affect the actual wording and intonation of the utterance. For pronouncing the utterance templates we use a Dutch Text-to-Speech system. The present navigation agent only mentions that he has performed his task ('Here is the information board.') and does not use templates.

2.5 Cross-media Reference Problems in the Virtual Environment

Because of its importance we spend a few words on the issue of cross-media references in interactions with our environment. This section presents problems we have to deal with in next versions of our environment. Presently we already have multiple agents, we have multiple input modalities and we have different media presentation possibilities available for output to the users. It will be clear that in such an environment any generated utterance (by users/visitors, individual agents, the system) can contain references to previous elements of discourse (words, phrases, utterances), that is, anaphora reference, or even to future elements of discourse, that is, cataphora reference. In our environment, discourse consists of a fusing of modalities and presentations.

For example, Karen may be asked a question using natural language keyboard input, she can answer with a combination of facial expressions, speech output and the presentation of a table with menu items. At this moment we have not bothered to fill our databases with pictures and previews of theatre performances. Therefore only as an example a single picture or video preview can be shown. However, in a full-fledged system containing not only reviews, but also pictures, interviews, video previews, etc., Karen should be able to present such information, either on request of the visitor (maybe after a suggestion of Karen that she can show it) or doing it as soon as becomes clear that a visitor might be interested (and the performance is not sold out already). Hence, in the interaction with Karen the visitor may not only make references to linguistic entities and concepts in Karen's utterances (and to the dialogue itself), but also to Karen's behavior, the way she looks, the things she is presenting on the screen, etc. When watching a preview video the visitor may ask: "Who is that girl?" Do we expect Karen to answer that question? Or simply: "Can I see the first video again?"

The user can use different input modalities to address agents. Menu items can be clicked on, natural language sentences can be typed, the viewpoint can be changed using mouse, speech or keyboard arrows, and in al these cases we can have implicit and explicit references to previous (and again, forthcoming) multimedia and multimodal elements of discourse, provided by user(s), agent(s) and system. Especially for our navigation agent it is clear that the visitor will make many references to entities that have not been part of earlier utterances but to actions that have been performed or to locations and objects that have been made visible for the visitor. At this moment we don't plan to include recognition of pointing gestures that are done by bringing hand and fingers close to a position on the screen. However, we are doing experiments with an eye-tracking system in order to detect where the visitor is looking at (in the experiments the goal is to activate one of several agents that are visible). This may at least help to disambiguate a question such as "What's this?". Similarly, when the visitor uses the mouse in order to mark or click a position on the screen in order to support such a question it becomes less difficult to give an expected interpretation to the question. Nevertheless, we may assume that there are situations in which the user asks "What's this?", feeling no need to give more information in order to get the question understood by our navigation agent. Hence, our navigation agent should not only, as is the case now, be able to obey speech commands which refer to elements of a pre-defined list of locations, but it should be given the intelligence to fuse different input modalities, to access stored knowledge about the environment and to reason about this knowledge in order to interpret and answer the questions that are asked by the visitor(s).

The agents and talking faces we are modeling are not only 'talking'. They display also limited nonverbal behavior (lip movements, facial expressions, gaze behavior, etc.), which can be referred to by other agents (in particular a visitor of the environment). See [4] for research results on cross-media references in dialogue.

3 VISUAL SPEECH, FACIAL ANIMATION, GESTURES AND MOVEMENTS

3.1 Introduction

The 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 to 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.

3.2 Facing the Information Agent

Figure 3 A Cartoon Face

We developed some virtual faces in a 3D-design environment. 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. Fig. 3). 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.

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 [3].

3.4 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.

4 TOWARDS 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? Why not look more closely at possibilities to be offered to:

- the professionals (directors, choreographers, stage crew, sound/light people, etc.)
- the performers, hence, the actors, the musicians, the dancers who present their work and prefer more or other interaction with each other or/and the audience

- the public in the role of audience attending a performance; not necessarily a passive audience, but also a web-audience that can (real-time) influence the running of things during a performance or can take part in a performance as an actor

We will not elaborate the possibility to use our environment for scenographic simulations. There are projects aiming at providing professionals tools and environments to help in pre-producing performances. Users can build a scenography of a performance, they can move through virtual models of stage sets in real time, they can experiment with lights or camera effects, change points of view, etc. In Rodriguez et al. [6] 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 [2]. 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.

In the traditional theatre, performers and audience are physically together. There is a focus of attention of the audience in things happening on stage and performers are aware of the audience's attention. Rather than to have one special physical space where performers and audience gather, now performers can be geographically dispersed and so can the audience. Moreover, there is no need to maintain the distinction between audience and performers. The environment should allow an (web) audience that can (real-time) influence the running of things during a performance or can even take part in a performance by taking the role of an actor. This requires special attention for the presence issue (see Reeve [5]).

In Sgouros [7] a distinction is made between direct interaction (a user joins an ongoing performance as one of the actors or players in it) and ritualistic interaction (applause, shouts, etc.). This latter form of interaction can influence the performance of an event (e.g., a football match) but it also unifies participants and mobilizes emotions and sentiment. The system that is described has a server which relays messages by the participants (players and spectators). The spectators can send messages to express emotions, to show approval and disapproval, to warn a player, etc., using some predefined and free-form messages with corresponding audio and image effects. This allows the server to detect audience preferences and to express reactions from the audience. Obviously, this may also allow the server to influence the event.

5 FUTURE RESEARCH & CONCLUSIONS

In this paper we reported about on-going research. 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. As may have become clear from the previous sections, our approach to designing a virtual environment for an interest community is bottom-up. In the beginning of 2000 a new project will started in which we get the opportunity to start working on a new version of our virtual environment using more fundamental approaches from software engineering themes such as design and specification and from agent technology.

REFERENCES

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

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

[3] 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.

[4] Proceedings "Referring Phenomena in a Multimedia Context and their Computational Treatment." ACL Workshop, Madrid, 1997.

[5] Reeve, C. Presence in Virtual Theatre. *BT Presence Workshop*, BT Labs, 1998.

[6] 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.

[7] Sgouros, N.M. Involving the audience in distributed, interactive entertainment systems. Proceedings EG MultiMedia'99 Workshop. *Media Convergence: Models, Technologies and Applications*. Milan, September 1999, Lecture Notes in Computer Science, Springer-Verlag, 113-120, to appear.