

# Multi-Media Access and Presentation in a Theatre Information Environment

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

## ABSTRACT

This paper discusses a virtual world for presenting multi-media information and for natural interactions with the environment to get access to this information. Apart from mouse and keyboard input, interactions take place using speech and language. It is shown how this 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. We investigate how we can increase the user's commitment to the environment and its agents by providing context and increasing the user's feeling of 'presence' in the environment. Moreover, we spend some notes on real-time and video performances in our virtual theatre.

**Keywords:** Virtual Reality, Multi-modal Access, Multimedia Presentation, Speech and Language Interactions, Agent Technology

## 1 INTRODUCTION

World Wide Web allows interactions and transactions through WebPages using speech and language, either by inanimate or live agents, image interpretation and generation, and, of course the more traditional ways of presenting explicitly pre-defined information by allowing users access to text, tables, figures, pictures, audio, animation and video. In a task- or domain-restricted way of interaction current technology allows the recognition and interpretation of rather natural speech and language in dialogues. However, rather than the current two-dimensional web-pages, the interesting parts of the Web will become three-dimensional, allowing the building of virtual worlds inhabited by interacting user and task agents and with which the user

can interact using different types of modalities, including speech and language interpretation and generation. Agents can work on behalf of users, hence, human computer interaction will make use of 'indirect management', rather than interacting through direct manipulation of data by users.

In this paper we discuss a virtual world for representing information and natural interactions about performances in an existing theatre. Apart from mouse and keyboard input interactions take place using speech and language. It is shown how this 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.

## 2 INFORMATION SPACES FOR INTEREST COMMUNITIES

Web-based digital cities have been around for some years. Like many computer games they have evolved from text environments to 2-dimensional graphical and 3D virtual environments with sounds, animation and video. Visitors, or maybe we should call them residents, of these cities visit libraries, museums, pubs, squares, etc., where they can get information, chat with others, etc. In these environments people get the feeling of being together, they are listening to each other and, in general, they take responsibility for the environment and theirs and others behavior in such environments.

Today there are examples of virtual spaces that are visited and inhabited by people sharing common interests. With virtual spaces or environments we want to refer to computer accessible environments where users (visitors, passers-by) can enter 3D environments,

browse (visual representations of) information and can communicate with objects or agents (maybe other visitors in the same environment). These spaces can for example, represent offices, shared workspaces, shops, class rooms, companies or cities. 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 or educational) interests, but share some common conditions (driving a car, being in hospital for some period, have the same therapy, belonging to the same political party, etc.).

As an example we mention a virtual world developed at a cancer research institute in Seattle. This world enables people struggling with cancer to obtain information and interact with others facing similar challenges. Patients, families and friends can enter the three-dimensional world (a rendering of the actual outpatient lobby), get 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.

### 3 A VIRTUAL THEATRE COMMUNITY: A CASE STUDY IN PROGRESS

#### 3.1 INTRODUCTION

We present research on visualization and interaction in a realistic model of an existing theatre. This existing 'Muziekcentrum' offers its visitors information about performances by means of a yearly brochure. In addition, it is possible to get information at an information desk in the theatre (during office hours), to get information by phone (by talking to a human or by using IVR). The database of the theater holds the information that is available at the beginning of the 'theatre season'. Our aim is to make this information more accessible by using multi-modal accessible multi-media web pages. A more general aim is to do research in the area of web-based services, in particular interactions in virtual environments using speech and language.

Our virtual theatre has been built according to design drawings of the architects of the building. Part has been realized by converting AutoCAD drawings to VRML97 (see Figure 1). Photographs and video recordings have been used to add 'textures' to walls, floors, etc. Sensor

nodes in the environment activate animations 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. They can enter the theatre and walk around, admire the paintings on the walls, enter the main performance hall, go to the balconies and, take a seat in order to get a view of the stage from that particular location. Information about today's performances is available on a screen that is automatically updated. Visitors may go to the information desk in the theatre, see previews and start a dialogue with an agent called 'Karen' (cf. Figure 2). Karen has her looks against her. We are working on a more attractive – not necessarily human-like - appearance for Karen. In 2000 we aim at user evaluation studies that will concentrate on questions about, among others, the need of reasonable realistic representations of the theatre information and transaction service interactions that are offered. Another aim that will be explored is the embedding of this particular virtual environment in a virtual cultural arena where people can ask, retrieve and explore information about theatre and music performances in general. Agents, like Karen, will help the users with these tasks.

#### 3.2 PROVIDING MODALITIES

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. Research done on information access and transaction in the context of modalities (and especially the sequential and simultaneous combination of modalities), that is, a multi-modality approach for WWW, can be embedded in the attempts to develop standards for access to the web and presentation on the web. For example, standards are being developed for speech access (voice browsing), 3D visualization (virtual reality modeling languages) and the combination of access and visualization (MPEG standards).

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. In addition to a multi-modality approach, there is 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.

### 3.3 PROVIDING VISUAL CONTEXT

We decided to visualize the environment in which people can get information about theatre performances, can make reservations and can talk to theatre employees and other visitors. VRML, agent technology, text-to-speech synthesis, talking faces, speech recognition, etc., became issues after taking this decision. They will be discussed in the next sections. 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.

It has become clear from several studies (cf. Friedman [3]) 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. They can be made to enjoy the interaction, they can be made to perform better, etc., all depending on the way the interface and the interaction strategy have been designed. It makes a difference to interact with a talking face display or with a text display. People behave differently in the presence of other people than they do when they are alone. 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 also 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 have concluded that apart from visualization in general, introducing talking faces and animated characters can help to make interaction more natural and to make shortcomings of the interaction technology, in particular speech and language, more acceptable to users. For that reason we are developing virtual faces and animated characters in our 3D-design environment. The faces consists of various three-dimensional coordinates and is connected through faces. These faces are shaded to visualize a three-dimensional virtual face. The 3D data is converted to VRML-data that can be used for real-time viewing of the 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 application. Some are rather realistic and some are more in a cartoon-style. Karen's face is the interface between the users of the virtual theatre and the theatre information system. A dialogue window is shown when users approach the information-desk while they are navigating in the virtual theatre.

Karen's face is capable of visualizing the speech synchronously to the speech output. This involves lip-movements according to a couple of visemes. The face visualizes facial expressions according to user's input or the system's output. For pronouncing the utterance templates we use the Fluent Dutch Text-to-Speech system which runs on top of the MBROLA diphone synthesizer.

## 4 AGENTS, INTERACTIONS & RETRIEVAL

### 4.1 A NAVIGATION AGENT

Clearly, the WWW-based virtual environment we are developing allows navigation input through keyboard and mouse. Such 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. In addition, a navigation agent has been developed that allows the user to explore the environment and to interact with objects in this environment by means of speech commands. Obviously, we do not want completely

separated modalities. It should be left to the user to choose between the interacting means or to use both, sequentially or simultaneously. A smooth integration of the pointing devices and speech in a virtual environment requires means to resolve deictic references that occur in the interaction, and the navigation agent should be able to reason (in a modest way) about the geometry of the world in which it moves. We slowly extend and improve the interaction and navigation intelligence of our present navigation agent. At this moment we are exploring the possibility of speech recognition for several clients on a central server and the advantages of making the navigation agent visible for the user. One of our conclusions is that current web technology hardly allows smooth integration of speech recognition and browsing a virtual world.

## 4.2 AN INFORMATION AND TRANSACTION AGENT

As mentioned before, a second agent called Karen allows a natural language dialogue with the system about performances, artists, dates, prices, etc. Karen wants to sell tickets. She is fed from a database that contains information about performances in some of our local theatres. Developing intelligence for Karen, in this particular environment, is a main aim of our project. 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 ([6]). In a prototype system we allow Karen to use a combination of speech synthesis and information presentation on the screen. 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. As mentioned earlier, in our experimental 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 and the user can click items to get more information. The virtual face that has been designed allows animation of lip and face movements and animation of some simple face expressions. For speech-image synchronization 3D images of visemes are called when corresponding phonemes are spoken. Since our application is web-based, it requires the solution of technical problems dealing with sending and compressing sound files, commands and synchronizing sounds and animations.

The theatre database which can be accessed by Karen (through the system's dialogue manager) contains information about performances. When is a particular performance, who are the performers, are there tickets available, how much do they cost, etc. A seat map allows the user to choose an available chair and to have a view of the stage from this particular chair. In order to decide on performances some previews (digitalized videos) are available that can be viewed on a monitor on Karen's information desk. Reviews of performances available in the theatre data base can be read aloud (using text-to-speech synthesis) by Karen.

## 4.3 MULTIMEDIA RETRIEVAL

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 present the 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 semantic network in which general knowledge about the domain has been stored. As an example, suppose someone asks: "Well, 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', 'actress', and maybe 'movie'. 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 semantic 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. These projects are

**Pop-Eye:** This project aims at automatic indexing of subtitles of movies and television programmes. This allows archives to be searched by keywords that appear in the subtitles and corresponding visual or soundtrack material to be retrieved.

**Olive:** The aim of this project is to develop a system which automatically produces indexes from the sound track of a programme (television or radio). This allows archives to be searched by keywords in the soundtracks and corresponding visual or soundtrack material to be retrieved.

#### **LIP: A language and image processing project**

Integrating 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 our main aim in the near future. 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 helping to have a nice evening.

## **4.4 FUTURE AGENTS: CONVERSATION, RETRIEVAL, FILTERING**

As may have become clear from the previous sections, our approach to designing a virtual environment for an interest community is bottom-up. We started with two agents and in order to introduce others a Java implemented agent platform has been designed. Using this platform other agents have been introduced, e.g., an information board showing upcoming events. Users can ask questions to the board and get answers. The board passes the questions to the other agents, in this case questions about performances are passed to the Karen agent.

At this moment 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 user. These agents can be passive that wait until they are put into action or they sense changes, take initiative and perform actions, e.g. to inform the user without being asked about new information.

One of our concerns in the near future will be the introduction of a conversational agent (which has some general knowledge about well known artists and some well known performances). In this way we have obtained three kinds of dialogues (information & transaction dialogues, command-like dialogues and conversational dialogues). A slight sharing of knowledge (in particular, preferences of the user) between agents will become possible. In a next step this knowledge of preferences should be exploited, not only in the interactions with the user, but also in designing an agent that retrieves information that matches with the users profile. Visualization of the domains accessible to users may help to guide the interpretation of questions and requests for retrieval (we give a different interpretation to a question about artists when we are in an opera building then when we are in a music hall).

## **5 THE FEEDBACK OF AGENTS**

## **6 THE ROLE OF PRESENCE**

'Presence', as defined by Lombard and Ditton [7], is the perceptual illusion of nonmediation, that is, 'a person fails to perceive or acknowledge the existence of a medium in his/her communication environment and responds as he/she would if the medium were not there.'

Causes and effects of presence have been studied in experiments, but the number of parameters involved, e.g., task and user characteristics, makes systematic research difficult.

From the previous sections it may have become clear that rather naturally emerging topics of our interest are closely related to the issue of presence. The environment that is offered and the locations that can be visited look familiar, the functions of several objects and what to do with them is clear from their appearance and the multimodality approach allows a variety of user input and the production of different sensory outputs. The agents in the environment are assumed to be friendly and cooperative and the embedding of talking faces and moving avatars in the environment will increase the tendency to treat agents as social actors, rather than as software modules.

However, increasing the user's feeling of presence has not been an explicit goal. Rather we have looked at possibilities to increase a user's commitment to the system (like we would in similar systems, e.g., for electronic commerce) with the aim to obtain co-operative behavior. One obvious reason which makes us loose a user is when use of clumsy technology (like speech and language technology) is not sufficiently backed up by context (including other modalities) which seduces the user to a certain interaction behavior and which helps to disambiguate the users utterances.

Needless to say that for many tasks and activities one may not expect that presence will necessarily enhance performance. On the contrary, presence can be noise which distracts a user from performing a task and which throws away the advantages of individual access from a home PC to, e.g., an information service. There is no need for a virtual waiting queue or the sound of a leaving plane when all we want to know is when the next flight to Amsterdam leaves.

It is clear that in our world mediated entities (our agents in the virtual world) play the role of social actors, they increase the feeling of presence and they help to increase the user's commitment to the system. A more technical reason to have agents as social actors is that they can influence the interaction behavior of the users in such a way that it stays in a certain domain of task and/or domain knowledge, hiding shortcomings of imperfect interaction technology, in particular speech and language technology. With other words, in order to increase the quality of the web-based information and transaction services we are offering, it seems to be useful to exploit the possibility to increase the role of social actors in our environment.

As already mentioned in Bates [2] (in the context of agents performing in virtual worlds), rather than build smart but narrow agents we can build agents with rather broad though perhaps more shallow capabilities. That is, for several applications we may be able to use agents

that are not intelligent, as long as they are not clearly stupid or unreal. That is,

*"An agent that keeps quiet may appear wise, while an agent that oversteps its abilities will probably destroy the suspension of disbelief. Thus, we propose to take advantage of the "Eliza effect", in which people see subtlety, understanding, and emotion in an agent as long as the agent does not actively destroy the illusion."*

It should be clear from previous sections that in our environment we decided to explore the possibility to increase the user's feeling of presence in order to increase his or hers commitment to the system. We certainly do not want to advocate such an approach in general. In our application we think this is a useful approach. We don't think this should necessarily be the general approach to the design of computer interfaces for multimedia retrieval.

## 7 THEATRE-RELATED PURPOSES OF THE VIRTUAL ENVIRONMENT

In the previous sections we concentrated on:

- the public, that is, men, women, children who want attend a certain performance or who want to know about performances in general, in a certain city or region, and at a certain date or in a certain period; the public has expectations about the information that is provided, it knows, for example, that different newspapers have different opinions about performances, hence, it is necessary to be careful in pushing the visitor to attend a certain performance
- the theatres, that is, the organizations that want to sell tickets, want positive reviews for the performances they hire, want to give correct and relevant information to the public, and have to offer contracts to the managers of artists and theatre companies in such a way that they are not loss-making

However, 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 that can be offered to:

- the professionals such as stage directors, choreographers, stage crew, sound and light people, etc.
- the performers, hence, the actors, the musicians, the dancers, the artists, authors and poets who

present their work and prefer more or other interaction with each other or/and the audience

- the public in its role of audience attending a performance; not necessarily a passive audience just enjoying a performance, but also a web-audience that can (real-time) influence the running of things during a performance or can even more explicitly take part in a performance by taking the role of 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. In these projects 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. An example of such an project is CATS (see Gil et al. [4]). Rather we would like to add the possibility to look at our environment as a stage on which we can have on-line performances or pre-recorded performances on request. It has been mentioned before, that the computer screen can be looked at as a stage and it has been argued that the theater metaphor can help in understanding human-computer interaction. The metaphor needs to be explored further, especially with regard to interface agents and the 'artistic' agents that have been introduced in CMU's Oz project (cf. Mateas [8]) and the Virtual Theatre project of Stanford University (see <http://www-ksl.stanford.edu/projects/-cait/publicity.html>). This requires further investigations in agent theory, but also in the possibilities to use WWW and the computer (not only its screen, but also data gloves, head-mounted devices, special clothing, etc.) to stage performances with real and virtual actors and (real) audience.

It is not unusual today to have meetings in virtual environments. Lectures have been given (and attended) in chat environments and meetings have been held in visualized meeting places. So, why not have live theatre performances over the web?

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, 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, both for actors and audience (see Reeve [11]).

Early online performances include a *Hamlet* parody on IRC (Internet Relay Chat) and *The Odyssey* by Homer. Well known is a VRML production of Shakespeare's *A Midsummer Night's Dream* performed live on April 26, 1998 (see Figure ??). The various avatars playing the roles were controlled by actors and the performance could be seen from any avatar's point of view of the stage (see <http://www.vrmldream.com/> for more details). In addition to the projects which concentrate on VR environments and interaction modalities there are projects on virtual interactive studio television (e.g., the European VISTA project with applications on Interactive Drama and Interactive Presenters; in this environment viewers can actively participate and direct the program being transmitted), on digital cinema (multi-treaded movies, interactive series with audience participation, etc.) and on agent-based TV broadcasting.

## REFERENCES

- [1] Agah, A. & K. Tanie. Multimedia Human-Computer Interaction for Presence and Exploration in a Telemuseum. *Presence: Teleoperators and Virtual Environments* 8 (1), February 1999, 104-111.
- [2] Bates, J. Virtual Reality, Art, and Entertainment. *Presence: Teleoperators and Virtual Environments* 1 (1), Winter 1992, 133-138.
- [3] Friedman, B. (ed.). *Human Values and the Design of Computer Technology*. CSLI Publications, Cambridge University Press, 1997.
- [4] Gil, F.M. et al. 3D Real-time graphic environments for theatrical and TV scenographic simulations. In: *Proceedings Nimes '98*, LLIA Nos 134-135-136, 1998, 163-167.
- [5] Laurel, B. *Computers as Theatre*. Addison-Wesley 1991; 2nd edition 1993.
- [6] 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.
- [7] Lombard, M. & T. Ditton. At the heart of it all: The concept of presence. *Journal of Mediated Communication* 3, Nr.2, September 1997.

- [8] Mateas, M. An Oz-centric review of interactive drama and believable agents. CMU-CS-97-156, Carnegie Mellon University, Pittsburgh, June 1997.
- [9] Nijholt, A., A. van Hessen & J. Hulstijn. Speech and Language Interaction in a (Virtual) Cultural Theatre. Proceedings *NLP and Industrial Applications*, Moncton, New Brunswick, August 1998, 176-182.
- [10] Reany, M. The Theatre of Virtual Reality: Designing Scenery in an Imaginary World. *Theatre Design and Technology*, Vol. XXIX, No.2, 1992, pp. 29-32.
- [11] Reeve, C. Presence in Virtual Theatre. *BT Presence Workshop*, BT Labs, 10-11 June 1998.