Virtual Cultural Environments: Exploration of Information

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

Abstract

There seem to be essential differences between shops, city halls, theatres and museums. However, when we try to design virtual environments allowing services, transactions, explorations and communication, are they really that different? Doesn't the 'virtual' in virtual environments allow us to introduce possibilities for visitors that extend those that are available in each of the mentioned physical environments and that have to be common for each of them? Multi-media presentation of information, multi-modal access to information, multi-media retrieval of information, filtering of information, helping users to explore and get access to information, multi-user communication and collaboration, etc.; it is not difficult to argue that these issues, not or only poorly available in physical environments, are useful and can be realized in virtual environments. We survey our research on these issues and our attempt to integrate this research in a virtual environment devoted to the presentation and exploration of theatre information. We try to make clear that the research topics we have to deal with do not differ from those where research explicitly deals with access to commercial environments, museum environments, educational environments and other service and transaction environments.

1 Introduction

There are lots of virtual reality websites where visitors can learn about archeological sites, monuments, Aztec temples, prehistoric caves, Stonehenge and historical events (e.g., battle grounds). Some recent research projects can be found in [Antonicelli *et al.*, 1999] and [Díez Higuera *et al.*, 1999]. Museums can put their collection on the Web, can turn web sites to learning sites or sites that can help students to do research. Visitors can get different options to experience a collection of a museum, to explore some historical building or to witness a historical event or urban development. An example of the latter can be found in the Nu.M.E project which aims at creating an interactive virtual environment that allows witnessing the evolution of the city of Bologna from 1000 to 2000 A.D. In this case – because of the time dimension - rather than 3D, the designers consider it a 4D environment (http://www.cineca.it/visit/NUME/). Particularly interesting and ambitious in this project are plans to design not only an online version, but also an immersive version of the environment. Similarly, in other contexts, one may think of presenting antiquities in VR simulated original settings.

In a virtual museum audience participation can take the form of cooperating in the design of an online exhibition, i.e., cooperating with the museum's curator, educator, exhibition and graphic designers by the intended audience (see e.g. [Lewis, 1997], for an exhibition for Illinois public school students). However, much more interesting is the possibility of audience participation in an exhibition during visiting and exploring VR web pages. Clearly, whenever the audience visits pages, they 'participate': they walk around, explore the environment or, e.g. in the Nu.M.E. project, see (pre-defined) changes in the environment. In addition, it is not at all unusual that the audience can click on items to activate events. Video fragments can be shown (see e.g. http://www.rijksmuseum.nl/ for the use of QuickTime VR in the Dutch Rijksmuseum) or audio fragments can be played (see and hear e.g. http://www.musikmuseet.se/ to experience the playing of the historical instruments that are displayed in the Stockholm virtual music museum). Well known, but not at all usual is the use of haptic devices, allowing a visitor to touch works of art or historic objects (teapots, vases, bronze figures, etc.) as is demonstrated in the Interactive Art Museum (University of Southern California, http://digimuse.usc.edu/haptics.html.

In this paper we will investigate audience participation in virtual environments using speech and language, allowing interaction with agents that inform, that can take care of transactions and that guide the visitor in the environment. Exploration of such an environment is driven by the curiosity of the visitor, rather than by predefined schemes. We present such an environment as a case study. Originally it was designed and used to inform visitors about theatre performances. However, we think that the ideas that have

* The research in this paper has been done in the context of the U-Wish project of the Dutch Telematics Institute, Enschede.

been realized and explored should not be restricted to this particular environment.

2 Exploring Similarities

Many distinctions can be made between shopping, going to a theatre performance, visiting an art exhibition in a gallery, going to a museum, buying a house and visiting a town hall to arrange a new passport or to registrate a newly-born child. These distinctions are useful and really make a difference. Nevertheless, we can also distinguish interactions of the people involved in these activities that is common rather than distinct. When the activity is not completely routine, people need help and advice. "What else is there to see?", "Do you have a cheaper products.", "Is there a Van Gogh painting from the same period?", "Why isn't this allowed?", "Do I have to pay for this?", etc.

We want to explore similarities rather than differences. Before doing that, it should be clear that different domains have different audiences; such audiences may require different interaction strategies and such audiences may require different types of help in order to get the information they want, to perform a transaction, to successfully conclude a negotiation or to be able to communicate with other users. Clearly, exceptions may lead to wonderful contacts between people and when we design information and transaction systems in virtual worlds, we certainly should take care that such contacts not only remain possible, but also are natural and even maybe more natural than in reality. Nevertheless, we should be aware of the fact that for the majority of people buying a pack of milk from a supermarket this act is extremely different from the act of discussing which performance, which art exhibition or which other kinds of cultural events are available for a visit this evening.

In this paper we take the point of view that despite such differences, there are so many similarities for which we need research results that, without neglecting the special needs for the domain of museums, the domain of theatres, the domain of commerce, the domain of (commercial and governmental) information services, etc., we first need to explore similarities rather than differences. Such similarities are:

- Visitors need to be able to explore an environment; this requires both visualization of an environment and the available information and the possibility to start dialogues about information, transactions and environment (available context);
- Visitors need to be able to change an environment; changes may differ from changing the number of available tickets for a performance, becoming a member of the noticeable audience (for a performance, an exhibition or a particular painting) to being able to influence the composition of an exhibition, the lightning during a performance, the behavior of actors during a performance, etc.
- Visitors need to be able to express themselves in their own, natural ways; the environment should be able to detect and interpret the different ways a visitor tries to communicate with the environment. That is, the environment has to understand multi-modal user input (keyboard, mouse, language, speech, gestures, touch, eye tracking, facial expressions, body movements, etc,).
- Visitors need help. This help may be provided by traditional means (e.g., context-sensitive pop-up windows), but it is more interesting to look at domain agents that know about the environment, that know about the information that is available, and that know about preferences of particular visitors.
- Visitors should be able to communicate with each other. That is, not only communication with domain agents, but communication with who ever shows interest in the same environment, the information it offers, but much more importantly, information that is not explicitly available in the environment, but that is available in the shared knowledge of domain agents, visitors and world wide web.

3 A Case Study: The Virtual Theatre

In this study we discuss a virtual world for presenting information about performances, associated artists and groups, availability of tickets, etc., for some existing theatres in our city. Traditionally, 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 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 IVR Technology) and to get information from local daily and weekly papers and monthly announcements issued by the theatre. The database of the theatre holds the information that is available at the beginning of the 'season'. Our aim was 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.

The dialogue system that has been embedded in this virtual environment is called SCHISMA [Lie *et al.*, 1998]. It is a (keyboard-driven) natural language accessible information system, able to inform users about theatre performances and to allow users to make reservations for performances. The system makes use of the database of performances in some of our local theatres. 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.

In general we do not really disagree with a view where users are expected to adapt to a system. If the system is attractive enough (or if there are no alternatives), they will. 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

Figure 1: Entrance of the Theatre

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 domain agents or with the help of other visitors in the environment.

In this way, the environment can develop into an interest community 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 is far from the actual situation in our environment, it must be possible that the environment offers the general audience the possibility to organize performances, meetings and to present (media) art.

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 suit their preferences for performing the 'task' that has to be done.

4 VR Embedded Interaction

4.1 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 (Fig. 1), 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 a 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 that is automatically updated using information from the database with performances. In addition, 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 'talking face' (Fig. 2).

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 perform-

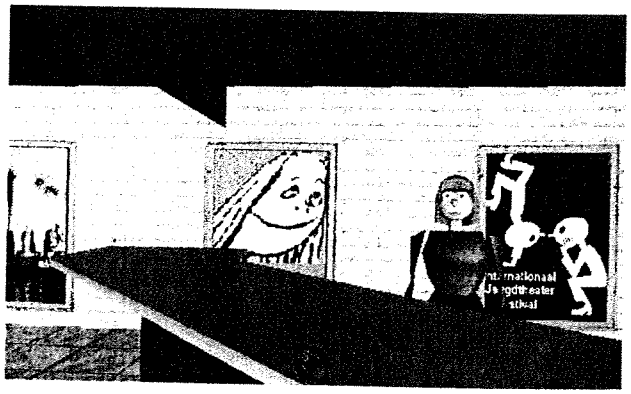

Figure 2: Karen at the Information Desk

[1] The system is accessible at http://parlevink.cs.utwente.nl.

ances do no have a video preview available yet, so we can not display them for every performance that is in the database.

4.2 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 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, 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 information systems will continue to increase. Dialogue systems wherein people can use normal wording become more and more common, but one of the problems in this kind of systems remains is the lack of context knowledge. 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.

4.3 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 WWW 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.

4.4 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, 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).

5 Agents in the Virtual Theatre

5.1 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 intelli-

gent 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 be done for other objects.

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.

5.2 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. In our current version of the dialogue system of which Karen is the face user utterances are simplified using a large 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, 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 performance. The user can click on items to get more information or can type in questions about items that are shown.

5.3 Navigating and Guiding

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 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 utterance. When it is successful, the agent guides the visitor to the location. When the 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.

5.4 Agents that Sell, Advise, Buy, . . .

Probably Karen's boss would like her to sell as many tickets as possible. A theatre director will have certain preferences in choosing performances for the theatre, but once 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 become inhabited by sales agents, hosts, guides, et cetera 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 establishing relationships with users through social "chit-chat" about family or work. That is, 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 need some superficial global knowledge that allows them to give socially appropriate answers to keep up a conversation about topics they hardly know of. They need some special hobbyhorses and some specific knowledge with which they can steer a conversation and which makes them 'believable' to the user. 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 the user should be able to communicate with other visitors of the environment or to address theatre agents who can share their specific knowledge (embedded in some kind of social conversation).

6 Visual Speech, Facial Animation, Gestures and Movements

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 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 demands of visitors or creators 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 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. For details of the design of the nonverbal response module of our agents see [Berk, 1998], [Nijholt and Hulstijn, 2000]. For gaze behavior of embodied agents see [Cassell and Thórisson, to appear].

7 Conclusions and Future Research

As may have become clear from the previous sections, our approach to designing a virtual environment for a particular domain has been bottom-up. At this moment the system has two embodied agents with different tasks and with simple interactions between them. Moreover, the agents do not employ a model of a user or of user groups. 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 order to do so there is a need to introduce an agent framework from which agents can be introduced in the environment in such a way that their intelligence, their believability and their ability to communicate follow from specifying the parameters of a model agent in this framework.

References

[Alberti *et al.*, 1999]. M.A. Alberti, G. Gallo & I. Jelinek (eds.). Proc. 20th Ann. Conf. European Association for Computer Graphics "Bringing to a New Life our Cultural Heritage", 1999.

[Antonicelli *et al.*, 1999] A. Antonicelli, G. Sciscio, R. Rosicarelli, G. Ausiello, T. Catarci & M. Ferrarini. Exploiting Pompei cultural heritage: the Plinius project. In [Alberti *et al.*], 100-103.

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

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

[Díez Higuera *et al.*, 1999] J.F. Díez Higuera, F.J. Díaz Pernas, J. Gómez García-Bermejo & R. Mompo Gomez. VRML-based system for a three-dimensional virtual museum on he web. In [Alberti *et al.*], 104-106.

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

[Lewis, 1997] Lewis, L.C. At Home in the Heartland Online: forming a museum/school resource via the World Wide Web. Museums and the Web Conference, Los Angeles, CA, 1997.

[Nijholt and Hulstijn, 2000] Nijholt, A. & J. Hulstijn. Multimodal Interactions with Agents in Virtual Worlds. In: *Future Directions for Intelligent Systems and Information Science*, N. Kasabov (ed.), Studies in Fuzziness & Soft Computing, Springer, 2000.