Maps, Agents and Dialogue for Exploring a Virtual World

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

ABSTRACT

In previous years we have been involved in several projects in which users (or visitors) had to find their way in information-rich virtual environments. 'Information-rich' means that the users do not know beforehand what is available in the environment, where to go in the environment to find the information and, moreover, users or visitors do not necessarily know exactly what they are looking for. Information-rich means also that the information may change during time. A second visit to the same environment will require different behavior of the visitor in order for him or her to obtain similar information than was available during a previous visit. In this paper we report about two projects and discuss our attempts to generalize from the different approaches and application domains to obtain a library of methods and tools to design and implement intelligent agents that inhabit virtual environments and where the agents support the navigation of the user/visitor.

Keywords: navigation assistance, agent technology, virtual environments, web-based services, multi-modal dialogues, usability and evaluation.

1. INTRODUCTION

In this paper we are concerned with the design and development of agents that help the user to find his way in web environments. We are concerned with helping and guiding the visitors of web environment to get the information he or she wants. In general this means that we need to guide the user in finding his or her way in popping up windows and menu's, clicking on menu items and links, etc. In our case, however, we have information presented in a virtual environment, the virtual environment itself is interesting for the user, the environment is inhabited by virtual agents and the user may want to know about these agents, the user may want to know what the system knows about his or her previous interactions, his or her interests, etc. Moreover, since the environment is a 3D virtual reality environment, the user is in fact a visitor who needs to know about this environment in order to find information or to perform a task. The question is: How can we assist the user/visitor when he or she is on our web pages displaying this 3D VR information?

In this paper we concentrate on two of our approaches to the development of a navigation assistant. The first approach is a rather general design of an agent that knows about the preferences of a visitor and, in addition, has some general knowledge of the environment that allows it to help the user to find his or her way in the environment. The general design should allow this agent to be added to different types of web environments. In the second approach we develop a navigation assistance agent that is able to engage into a natural language dialogue with the user about its task. We are far away from a situation where an agent has real-life knowledge about its relation with the environment, its task and the user it tries to assist. Nevertheless, in our projects we try to model such aspects in a navigation-supporting agent. Both approaches are part of our so-called Aveiro (Agents in Virtual Environments) Project. In this project we have a comprehensive, agent-oriented design approach to virtual agent-inhabited environments. Agents are ready to inform the visitor about performances, artists, reservations and the environment itself as it is displayed to the visitors.

Before going to the details of our navigation agents we give a short description of this virtual environment. The environment is the virtual equivalent of an existing theatre in our hometown. The theatre has different floors, a main performance hall, a lounge, stairs, etc. A receptionist in the form of an embodied agent and called Karin is available to answer questions about performances, performers, available seats and can make reservations. Questions can be asked using the keyboard and natural language. The receptionist has a database available with the actual theatre performances for the current year. A text-to-speech synthesis system is used to mouth her answers to the visitor. The environment has been built using VRML. Visitors can walk around in the environment, visit the different locations and the receptionist. In Nijholt & Hulstijn [10] a rather comprehensive survey of the environment is given.

From the visitor's point of view the need of an other agent emerged. To whom do we address our questions about the environment itself? To whom do we address our questions about how to continue, where to find other visitors or where to find domain-related information? At this moment we are following different approaches to solve this problem. The approaches are related and can be integrated since all of them are agent-oriented and are oriented towards a common framework of communicating agents. In addition, we build this framework in such a way that different agents with different abilities can become part of it: a simple animated piano player, a baroque dancer that 'understands' the music she is dancing on, Karin who knows about theatre performances and our navigation agent who knows about the geography of the building.

The research we report about in this paper has partly been done in the context of the U-WISH[1] project [8]. This project aims at

[1] In U-WISH (Usability of Web-based Information Services for Hypermedia), several partners of the Dutch Telematics Institute

developing Web-usability knowledge and methods to improve the accessibility of Web-based information services for hypermedia. This is done in particular by developing design principles for navigation support, addressing and developing techniques and tools for enhancing the accessibility and applying and assessing the principles, techniques and tools. Aspects and progress of this project have been discussed in several of our papers. For example, an overview of our research done on developing and evaluating specification techniques that are not only suited for describing interactive web-based services, but also can serve as a means of communication between designers and developers of such systems, has been published in Nijholt et al. [9].

2. NAVIGATION ASSISTANCE: AN INTRODUCTION

A well known definition of navigation is from Darken & Sibert [2]: "Navigation is the process by which people control their movement using environmental cues and artificial aids such as maps so that they can achieve their goals without getting lost." Others have added to this definition, e.g., by taking into account domain characteristics, knowledge the navigator has, the goals of the navigator, possible cues and aids, etc. (Krieg-Bruckner et al. [6], Werner et al. [16], Spence [13], Vinson [14], Volbracht & Domik [15], and Nash et al. [7]). In [3] Darken and Sibert suggest some design principles for navigating in virtual worlds. These principles concern, among others, the organization of the environment. They advise to divide large-scale worlds into distinct parts that are simply organized. Structure should be provided that enables the visitors to mentally organize the environment. Directional cues and visual and auditory cues can be added to the environment in order to ease recognition by the visitor of these parts and subparts and their organization. Other principles are concerned with the addition of map-like information. They are intended to present spatial information directly in such a way that the visitor can produce a flexible orientation-independent representation of the environment. The basic principles are to show paths, landmarks, subparts and their organization and the user's position. Furthermore they advise to orient the map towards the user such that "the forward-up equivalence principle is accommodated": the map is presented in the same orientation as the environment itself so that for users it feels as if the map was in front of their chest.

The principles that are mentioned are indeed derived from real world navigation support. However, especially in virtual worlds more real world navigation principles can be included. For example, in the real world we can ask someone we accidentally meet whether we are on the right way. Moreover, in some situations we can turn to special people that are here to take care of us when we have trouble to find our way. There are also situations where we can observe what others are doing or what others have been done in the past because they have left tracks.

A virtual reality environment can mimic a real world environment. We can indeed add agents to a virtual environment that know about the environment and that can interact with a visitor about what this visitor wants to achieve, where he wants to go, etc. In addition, virtual worlds can be multi-user worlds. Like any other virtual world, many other people can have visited them in the past. These people may have left traces. It is quite possible to make this visible in virtual environments and visitors may decide to explore an environment by following those traces. However, in a multi-user environment other users can provide real-time help or their behavior can be observed and imitated. These are examples of social navigation (Benyon & Höök [1], Höök [5]).

In our environment we have not (yet) implemented the possibility of social navigation, although some routes through the environment can be made more likely than others, e.g., by showing walking lines on the floor based on frequencies in the logs we collected over the years. We have also hardly experience with the behavior of multiple visitors in our environment. This happens very infrequently, to be honest, mostly during demonstrations only. However, adding an agent to our environment that knows about the geography of the environment and that knows how to interact about these matters with a user, is a rather natural extension of our present environment because of the presence of avatars representing visitors, animated 'characters' and Karin, the receptionist agent.

It is useful to mention some design criteria that can be considered when introducing a navigation agent. For example, will it have a more or less human-like appearance, that is, will it be embodied or will it appear in the form of a window with possibly some menus? An embodied agent will tend to be more obtrusive (or less unobtrusive) than a non-embodied agent. When embodied, its appearance may be a cartoon personality, a human-like personality, or another character. Whatever the character that is decided upon, additional choices may concern among others the agent's facial expressions, its nonverbal and verbal responses to a user's communication and its other, body animations. An animated agent has particular competence. It can show how to manipulate objects and it can employ gesture to focus attention. Various examples of embodied agents that provide help when visiting commercial company web pages already exist and are being used, despite diverging opinions in the scientific community about the effects of embodiment on the user (see e.g. Dehn & van Mulken [4], Rickenberg and Reeves [12] and our introduction to a panel discussion on this topic in [11]).

There are also design criteria when we consider the behavior of the agent with respect to its context, including the visitor. A pro-active agent may address a user without being asked for that by the user, while a reactive agent has to be deliberately activated by the user, for instance by means of a question or a mouse click. Note that this is a distinction that refers to user perceptions. From a technical viewpoint nearly all agents will be reactive, acting either in response to a question or to some observed user behavior. For example, an agent that offers guidance when a user seems to 'walk in circles' is acting reactively from a technical point of view, but a user will not consider this behavior as reactive, probably. To be complete, it is added here that agents can be designed to be pro-active as well as reactive, addressing users without being asked for that but also reacting when users ask for assistance.

Whereas these latter criteria concerned agent characteristics that users may readily observe, a third criterion can be added that pertains to a characteristic less obvious for users: the agent's adaptiveness, which we define as the agent's ability to include knowledge about user-related changes in the assistance offered to that particular user.

---

participate: the knowledge institutes TNO, CTIT and CWI, and the companies KPN, Océ, and Rabofacet.

To provide assistance, an agent must have adequate, up-to-date knowledge of the virtual environment. This environment will change constantly. In the virtual theatre, for instance, new performances are added to the schedule, and yesterday's performances are removed from the schedule. Moreover, performances can be fully booked, they can be cancelled, or the starting time can be changed. A virtual agent whose task is to provide reservation assistance needs to have the most recent information about the performances schedule. This agent is not adaptive in the sense of applying user-related information. If an agent takes user information into account that was collected at some point in time - for instance, at the occasion of the user's first visit or registration - and that information is not a subject of later verification, revision, or extension, there is no user-related dynamics. In that case we choose not to consider the agent as adaptive but rather use the term personalized agent.

Figure 1: A 2D map of the theatre displayed in the virtual world

To be able to offer user-specific assistance, an adaptive agent will have to make use of information that relates to a particular user. It is clear that in many situations we can expect different user interaction behaviors and different user preferences with respect to the "content" that is offered. These differences follow from different interests, background, culture, intelligence and interaction capabilities of users. These issues can become part of a user profile and will be used by the agents to anticipate the users preferences or to guide a user in reaching a particular place of the user's choice. In our research we have not yet paid attention to this kind of adaptiveness.

3. VR NAVIGATION SUPPORT IN U-WISH

In the context of the U-WISH project several concepts and tools for navigation have been introduced and developed for our virtual environment. One of them is the categorizing landmark. These landmarks are color-coded backgrounds that show the user relationships between web pages. This is a general principle, which we also added to our virtual reality environment and the navigation assistant window. An other concept is a 2D map of the environment displayed, on request, over the virtual world. The third concept, the navigation assistant, provides the user with information about possibly interesting pages (or information about objects and locations in our virtual environment), based on some (fixed and very limited) information about the user.

The categorizing landmarks belong to the principles and cues that we mentioned in section 2 and that concern the structuring of the environment in subparts. The margins of the browse window and the navigation window (see below) take the color of the subpart of the environment when the user enters that part of the world. The colors are also visible in the map that has been implemented. As mentioned, it appears on request, the user's position is shown and the map is turned according to the orientation of the user according to the forward-up equivalence principle. In Figure 1 a view on the environment with the map displayed over it is presented.

As mentioned, the navigation assistant described in this section was developed as a general tool for web-based services in general. Therefore it has been developed and introduced in two different environments, a traditional web page environment providing information about one of our partners in the project and the web sites displaying our virtual theatre environment.

This (non-embodied) agent is pro-active: it presents advice in a (small unobtrusive) window without being asked to do so by the visitor. In the future, it will also be possible to ask this agent to give advice or to make suggestions. The agent is adaptive in the sense that the visitor's (changed) position is used to give advice. Adaptiveness in the sense of using dynamically changing visitor profiles is not planned to be added in the near future.

The navigation assistant knows about the visitor's profile. Currently, in our experiments, this profile contains information about interests, profession and means of transport, distinguishing a very limited number of different visitors. We have not yet research going on obtaining and adapting visitor profiles, the current profiles have only been introduced to be able to set up experiments with users. Obviously, the navigation assistant knows about the web-site (what is displayed of the virtual world) and it has some knowledge of the interaction that has been going on with the user, that is, it knows about previous suggestions to the visitor. These three knowledge sources (visitor profile, web-site and interaction history), together with the current position of the visitor, are available to the assistant when it makes suggestions to the user about what to do next.

Especially in a virtual environment the current position and the orientation of the visitor is important. The virtual environment has (invisible) sensors that are triggered when the visitor comes close. In this way, if, for example, the visitor passes a poster announcing a particular performance that suits the interests of the visitor, the navigation assistant can draw the attention of the visitor to this particular poster. This is done in a modest navigation window right below the browser of the virtual world. It is also possible to present hyperlinks to the user in this separate window that can be clicked by the visitor. In this version all the initiative is with the assistant. Obviously, the user is free to follow the assistant's advice or neglect it.

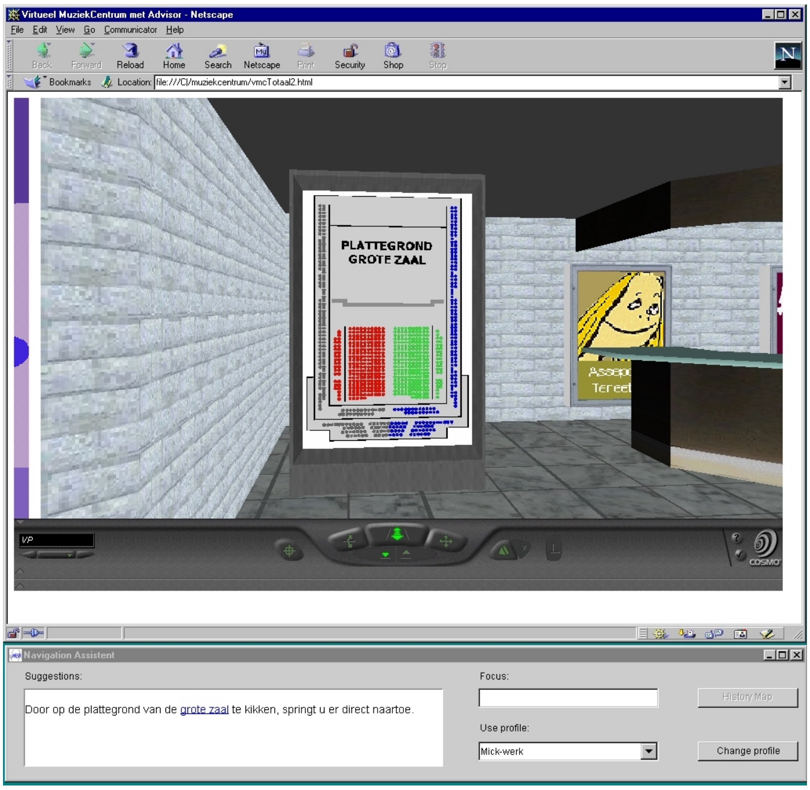

Figure 2: Browser and navigation agent window

In Figure 2 we have displayed the navigation window together with the browser window. In this case the visitor is informed that by clicking on the seat map, which is visible in the browser window, he will be teleported to the main performance hall.

During browsing, the visitor has the possibility to temporary adjust the visitor profile. This can be done by entering an additional keyword in the window of the navigation assistant leading to more focused interests in the profile while visiting web pages. This allows the assistant to give better suggestions. The design of this assistant is generic, it can be considered as a site-independent plug-in. This is certainly not the case for the assistant we discuss in a next section of this paper. However, we think that both approaches can be integrated into one design for future applications.

4. OBSERVATIONS ON AN AGENT FRAMEWORK

For our virtual music center (VMC) an agent-based architecture is under construction in which all sorts of *agents* could be added successively in order to improve an assistant's advice step by step. An example configuration of agents, together with a few non-agent processes, is shown in Figure 3. The main advisor agent is built up internally from a co-coordinating 'super agent', as well as specialized agents, like a position tracker, a word spotter, etc. Each of these specialized agents focuses on a particular aspect of user behavior, and provides its own advice. The co-coordinator agent filters and combines these, and sends them to agents that interact directly with the user of the system. This approach enables the exchange of information with 'closed' objects (Java applets, VRML viewers, Flash, etc) that web-sites can contain. In order to do so, we insert mediating agents like the eavesdrop agent that translate between a non-agent processes and our agent based architecture. Most of the agents do not deal with Internet connections and communication directly. Instead, communication with other agents is delegated to specialized communicator agents, not unlike the facilitators from other agent platforms. As a result, agents can be allocated, either on the machine that runs the virtual music center user interface, or on a remote server, whatever turns out to be best.

5. THE NATURAL LANGUAGE NAVIGATION ASSISTANT

In order to investigate the problems and solutions of communicating in natural language with a navigation agent in a virtual reality environment we introduced a version of our environment where we have added a window to the virtual reality browser which displays a detailed floor map with positions of different objects and locations and also possible routes between them. This has become a project on its own and the reader should not confuse the description here with the one in section 3.

Associated with the map a natural language accessible navigation agent was introduced. This work is in progress, meaning that this special version of the system is there, but no effort has been done to add 'graphic sugar' to the layout and the integration of the different windows that are used. When user experiments start this has to be worked on first. In Figure 3 we display the current floor map and the agent window of this version of the system.

The visitor can ask questions, give commands and provide

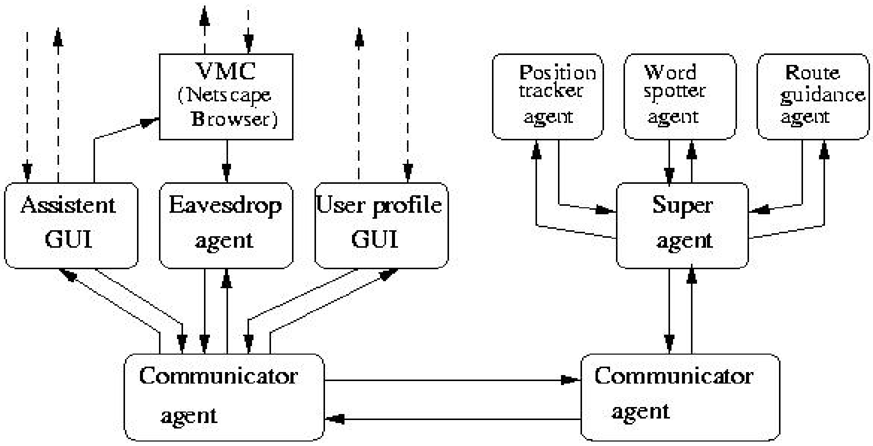

Figure 3: The communication infrastructure of the VMC

information when prompted by this navigation agent. This is done by typing natural language utterances and by moving the mouse pointer over the map to locations and objects the user is interested in. On the map the user can find the performance halls, the lounges and bars, selling points, information desks and other interesting locations and objects. The current position of the visitor in the virtual environment is marked on the map. While moving in virtual reality the visitor can check her position on this floor map. When using the mouse to point at a position on the map references can be made by user (in natural language) and system to the object or location pointed at.

As mentioned, this navigation agent has to be accessed by natural language. We have annotated a small corpus of example user utterances that appear in navigation dialogues. On the one hand we have complete questions and commands. On the other hand we have also short phrases that are given by the user in reply to a clarifying question of the navigation agent. An example of a question is: “What is this?” while pointing at an object on the map, or “Is there an entrance for wheel chairs?” Examples of commands are “Bring me there.” or “Bring me to the information desk.” Examples of short phrases are “No, that one.” or “Karin.” From the annotated corpus a grammar was induced and a unification-type parser for Dutch can be used to parse these utterances into feature structures.

Three agents communicate to fill in missing information in the feature structure (when the information given by the user in question, answer or command is not yet complete) and to determine the action that has to be undertaken (answering the question, prompting for clarification or missing information, displaying a route on the map or guiding the user in virtual reality to a certain position). This is done by the navigation agent in co-operation with the dialogue manager and the Cosmo Agent. The latter can ‘talk’ to the Cosmo Browser using its EAI (External Authoring Interface) to retrieve the current position of the visitor. Not yet implemented is the possibility that not only the position but also what is in the eyesight of the visitor in virtual reality can be retrieved. This will allow more correct reference solving in the dialogue.

The natural language interaction between the navigation agent and the user allows the user to play an active role in the process of navigation. The navigation agent is reactive: the visitor can ask about existing locations in the theatre. The user can type a question like "Where do I find the coffee bar?" or a command like "Bring me to the coffee bar, please" and the system can react by answering the question in two ways: it can indicate the place on a map, or it can navigate the visitor’s viewpoint through the environment along a route to this destination. In order to do so the agent needs to know:

- how objects in the inventory of the environment are referred to by means of a natural language expression ("the coffee bar")
- how the actions it can perform can be referred to by means of natural language ("bring").
- what communicative act the user is performing by his utterance (is the user asking for information, or asking the system to do something)

Because the visitors will be aware of the visual context, in natural language interaction they will probably use references to that context. Hence natural language understanding cannot be seen as an isolated activity that is carried out by some language processing module that is independent of the virtual environment. Rather, the interpretation of natural language sentences is coupled to what is seen in the virtual world at the moment the sentence is uttered by the user. For instance, our advisor might suggest going through ``the door", in case exactly one door is visible. The use of words like ’this’, ‘that’, ‘there’, ‘here’ (deictic references) can only be understood by a natural language capable agent if this agent is able to recognize what is in the neighborhood of the user, or what can be seen by the user. Also the agent should be able to recognize objects that have recently been referred to in the dialogue and that could have been used in the utterance at that particular position. Such objects are stored in a focus list. We illustrate this by an example dialogue:

- User action: "Where do I find the coffee bar?"
- System action: shows the coffee bar on a map
- User action: "Please, bring me there."
- System action: navigates to the coffee bar.

Since the system has been able to solve the coffee bar reference, and stored the information in the focus list, it can attach the indexical "there" to the object referred to earlier in the dialogue. If the user asked the way to the coffee bar and then tries to find his way through the environment, the navigation agent should remember what the user is looking for so he can interrupt if he notices that the user navigates in a wrong direction: "you should go left here, if you look for the coffee bar". In case the reference problem could not be solved, the system can ask for more information. When the visitor’s utterance is about performances, the navigation agent may attempt to contact Karin, the information and transaction agent.

6. CONCLUSIONS

The prototype navigation agents that we discussed here are certainly not our final solution in assisting visitors of our virtual environment. In the next phase of research we need to concentrate on the communication with other agents that are available in the virtual theatre environment. How can we take care that a visitor’s question reaches the appropriate agent? How can we model the history of interaction in such a way that different agents do not only know about their own role in this interaction but also about others. What should agents know about themselves, the visitors and the environment?

The agents that currently inhabit our virtual world do not comply with ‘standards’ that have been developed in (multi-) agent frameworks. Our ‘Communicator’ (see section 4), however, provides the basis for an agent communication framework on which we can build further. Recent research aims at introducing BDI (Beliefs, Desires, Intentions) like agents in this framework. This will make it necessary to redesign existing agents, but the advantage is of course that intelligent behavior can be designed more uniformly, including the ability of such an agent to interpret and generate natural language utterances and to have coordinated nonverbal communication between agents and visitors.

Obviously, since we mentioned nonverbal communication, one of the things we should consider for the future is whether we will make a navigation assistant an animated character. This can be a face displayed at a fixed position in the environment, or in a separate window, using speech synthesis, lip synchronization and some facial expressions, or even a character that walks through the environment with the user and that is able to point to objects or show directions.

Acknowledgements. We are grateful to Jeroen van Luin and Rieks op den Akker for their contributions to section 5 of this paper.

7. REFERENCES

[1] D. Benyon & K. Höök. Navigation in information spaces: Supporting the individual. In: S. Howard, J. Hammond & G. Lindegaard (eds.), *Human-Computer Interaction: INTERACT'97*, Chapman & Hall, 1997, 39-46.

[2] R.P. Darken & J.L. Sibert. A toolset for navigation in virtual environments. In: *Proceedings of UIST '93*, Atlanta GA, ACM, 1993, 157-165.

[3] R.P. Darken & J.L. Sibert. Way finding strategies and behaviors in large virtual worlds. In: *Proc. of CHI '96*, ACM, 1996, 142-149.

[4] D.M. Dehn & S. van Mulken. The impact of animated interface agents: a review of empirical research. *Int. J. Human-Computer Studies, 52*, 2000, 1-22.

[5] K. Höök et al. Towards a framework for design and evaluation of navigation in electronic spaces. *Persona Deliverable* for the EC, 1998.

[6] B. Krieg-Brückner, T. Röfer, H.O. Carmesin & R. Müller. A taxonomy of spatial knowledge for navigation and its application to the Bremen autonomous wheelchair. In: C. Freksa, C. Habel, & K.F. Wender (eds.), *Spatial Cognition, Lecture Notes in Artificial Intelligence 1404*, Berlin: Springer Verlag, 1998, 373-397.

[7] E.B. Nash, G.W. Edwards, J.A. Thompson & W. Barfield. A review of presence and performance in virtual environments. *Int. J. Human-Computer Interaction*, 12, 2000, 1-41.

[8] M.A. Neerincx, S. Pemberton and J. Lindenberg. U-WISH; Web usability: methods, guidelines and support interfaces. TNO-report TM-99-D005, TNO Human Factors Research Institute, 1999.

[9] A. Nijholt, O. Donk & B. van Dijk. Specification Techniques for Multi-Modal Dialogues in the U-Wish Project. Proc. *Systemics, Cybernetics and Informatics (SCI'00)*, Vol. II: Information Systems Development. B. Sanchez, N. Nada, A. Rashid, T. Arndt & M. Sanchez (eds.), Orlando, USA, 2000, 340-345.

[10] A. Nijholt & J. Hulstijn. Multimodal interactions with agents in virtual worlds. In *Future Directions for Intelligent Information Systems and Information Science*, N. Kasabov (ed.), Physica-Verlag: Studies in Fuzziness and Soft Computing, 2000, 148-173.

[11] A. Nijholt. Agents, believability and embodiment in advanced learning environments: Introduction to a panel discussion. In: Proc. IEEE *International Conference on Advanced Learning Technologies (ICALT 2001)*, Madison, USA,2001, to appear.

[12] R. Rickenberg & B. Reeves. The effects of animated characters on anxiety, task performance, and evaluations of user interfaces. In *Proc. of CHI'00: Human Factors in Computing Systems*, The Hague, The Netherlands, 2000, 49-56.

[13] R. Spence. A framework for navigation. *Intern. J. Human-Computer Studies*, 51, 1999, 919-945.

[14] N. Vinson. Design guidelines for landmarks to support navigation in virtual environments. In: Proceedings of *CHI '99*, ACM, 1999, 278-285.

[15] V. Volbracht & G. Domik. Developing effective navigation techniques in virtual 3d environments. In: J.D. Mulder, & R. van Liere (eds.), *Virtual Environments 2000*: Proc. of the Eurographics Workshop, Springer, 2000, 55-64.

[16] S. Werner, B. Krieg-Brückner & T. Herrmann. Modeling navigational knowledge by route graphs. In: C. Freksa, W. Brauer, & K.F. Wender (eds.), *Spatial Cognition II*, LNAI 1849, Berlin: Springer Verlag, 2000, 295-316.

[17] J. Zwiers, B. van Dijk, A. Nijholt & R. op den Akker. Design issues for intelligent navigation and assistance agents in virtual environments. Proc. of the TWLT 17 Workshop on Language Technology: *Learning to Behave. Workshop I: Interacting Agents*, Enschede, The Netherlands, 119-132.