VIRTUAL PRESENTERS: TOWARDS INTERACTIVE VIRTUAL PRESENTATIONS

Anton Nijholt
Human Media Interaction Group
University of Twente
Enschede, the Netherlands
anijholt@cs.utwente.nl

Abstract – We discuss having virtual presenters in virtual environments that present information to visitors of these environments. Some current research is surveyed and we will look in particular to our research in the context of a virtual meeting room where a virtual presenter uses speech, gestures, pointing and sheets to give a presentation generated from a script.

INTRODUCTION

Information has to be presented. It should be possible to interact with media that allow the presentation of information. There are many ways to present information. When we ask directions on the street we can get quite detailed information, about how to go from the current location to a desired location. The explanation consists of verbal and nonverbal utterances, that is, sometimes the verbal utterances support pointing gestures or gestures that explain objects (landmarks) and situations that will be met, sometimes the gestures support the verbal utterances. A museum guide verbally and nonverbally (using gestures and gaze) interacts with his or her audience to explain the interesting parts of a sculpture or painting, addressing one, several or all persons in his or her audience. And, after having explained a piece of art, the guide will make it clear where to go or look next, again by verbal and nonverbal means of addressing the audience. Yet another way of presenting information is to give a presentation, using overhead, data, or video projector. Clearly, also in these cases we can expect that the presenter will use deictic references to pictures, bullets, and texts fragments that appear during a presentation, e.g., a power-point presentation, on the screen. In this paper we survey our research on presenting, where presenting means that - verbally and nonverbally - explanations are added by a virtual human-like presenter, to a scene or an object that is visible to this presenter's audience. The object may be a power point sheet displayed on a screen, but it can as well be a painting, displayed and explained in a virtual museum environment.

There are many examples of research efforts on automating presentations. Sometimes they aim at supporting the human presenter, sometimes they aim at giving the presentation task to a robot or virtual agent. In the latter case we can have one or more virtual embodied agents added to 2D web pages or power point slides on which the presentation takes place, but we can also have the situation where a lecture, meeting or exposition room has been virtually represented and has been provided with a virtual presenter, guide or assistant that can present and explain to visitors. Whether the efforts are aimed at supporting a human presenter, a robot presenter or a virtual presenter, knowledge about the contents of the presentation and knowledge about the interaction between presenter and audience need to be modeled. These different types of artificial presenters all should make use of the same kind of knowledge.

BACKGROUND OF OUR RESEARCH

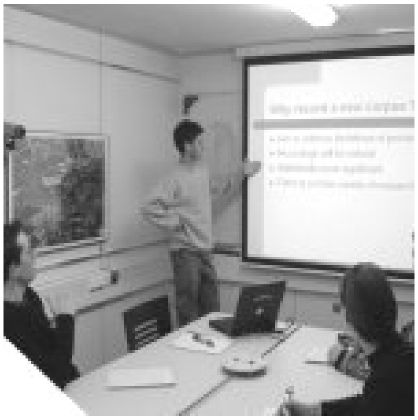

Figure 1. AMI meeting presentation

Our research on presenting is – in principle - not tuned to one particular application. However, it is part of research in a FP6 European IP on Augmented Multi-party Interaction (AMI) [1]. This project is concerned with the modeling of interactions in a smart meeting environment, with the aim to provide real-time support to the meeting partners and to allow off-line multimedia retrieval and multimedia browsing of information obtained from a particular meeting. For that reason our examples are drawn from the domain of meetings. However, as explained in [5], the technology and the models that are being developed and designed can find their way in all kinds of applications of smart environments, including museum and exhibition environments that give real-time support to their visitors. The technology that is being developed allows to detect, track and identify people in a particular environment and to interpret their activities and their interaction with other people or with objects and locations in the environment. Our main line of research is to translate our findings on meeting modeling into tools (meeting assistants, meeting browsers, meeting visualization, etc.) that can be used in real-time during a meeting or offline to browse through a previous meeting. Part of this research is how to model presentations during meetings (cf. Figure 1) and how to translate them to sufficiently realistic presentations in a virtual reality representation.

A CORPUS OF PRESENTATIONS

In the AMI project a corpus of meetings is being collected. These meetings contain whiteboard and PowerPoint presentations. Especially in this latter case, we have a presenter explaining what is already visible on the screen, not essentially different (apart from the content) from a guide in a museum that explains a painting or a sculpture to a group of tourists visiting the museum. Gestures are made, there is pointing to the interesting parts, and there is some interaction with the audience, verbally and nonverbally. The corpus that was available at the start of the AMI project consisted of mock-up meetings, where, for example, during a meeting someone stands up to deliver a presentation. During the project new corpora will emerge, depending on the research interests of the different partners in this large-scale European project. One, fully unstructured corpus that has been added is a series of thirty videos of presentations during a workshop associated with the project. Designing models for multimodal human presentation goes together with designing annotation schemes and annotating presentations from the corpora. For this purpose some annotation tools have already been developed that make it possible to relate spoken content with gestures. New annotation tools are developed that also take into account pointing and other gestures that refer to parts of a scene, e.g. part of a painting, a sheet of a power point presentation or a location.

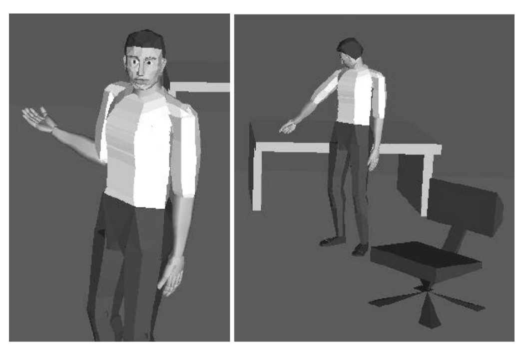

Figure 2. Noma & Badler's virtual presenter

TOWARDS VIRTUAL PRESENTERS: RELATED RESEARCH

How can we put knowledge about presentations in use? That is, how can we model this knowledge in software and hardware and then give computer support in situations where presentations need to be delivered? Our aim is to have virtual presenters available on websites, 3D and virtual reality environments that are human-like (i.e., embodied conversational agents) and in which this presentation knowledge is modeled. Examples of such presenters have already been made available [4,10], either in 2D or 3D form. Our work is related to that of Noma and Badler [7]. Their presenter (see Figure 2) can make presentations in a 3D virtual environment or on the WWW. It gets its input from speech texts with embedded commands that relate to the presenter's body language. This presenter behaves as a TV presenter, e.g., a weather report presenter that knows about the camera and the presentation screen, but does not have an audience in its direct environment.

Much more related research is going on. Let us first look at project very much related to our AMI project. Like this project, CHIL - Computers in the Human Interaction Loop - is an Integrated Project on multimodality under the European Commission's 6th Framework Programme. Both projects cover the EU attempts to stimulate research on multimodality. While AMI concentrates on meetings (where we may have short presentations by the meeting participants), CHIL research gives more attention to support that can be given to lecturers and their audiences. Hence, in CHIL they look at human interactions in lectures and one of the research aims is to develop a so-called Attention Cockpit, an agent that tracks the attention of the audience and provides feedback to a lecturer. In this way the lecturer can be notified when the audience becomes disconnected or when a participant intends to ask a question. Potential services for the speaker that are foreseen are among others Audience Monitoring and Info (provide information about the background knowledge of the audience to the speaker and monitoring of audience eye contact, facial expressions, body language and questions) and Slide Changing (change slides automatically based on monitoring both the speakers speech and questions asked) [11]. Other potential services take care of microphone control and informing the speaker about the quality of speech. Obviously, although CHIL is not concerned with designing a virtual presenter, research results for the Attention Cockpit help to model capabilities of the virtual presenter in our research. Presently CHIL research is in progress. One supporting action has been a survey conducted to collect data from people attending presentations and people giving presentations [11]. On short notice we can not expect comprehensive modeling of presenter-audience interaction in automated presentations, but it is certainly something that is being addressed in the ambitious AMI and CHIL projects.

It is much more simple, and more closely to state-of-the-art research, to concentrate on the presentation contents and the relation between the contents and the artificial presenter, leaving out any possible interaction with an audience. This brings us on research in the area of scripting and mark-up languages. An example is the Multimodal Presentation Mark-up Language (MPML) [2] which also allows affective control of the presentation by incorporating models for emotion, mood and personality for the artificial presenter. The language has also been used in combination with a humanoid robot rather than with a virtual agent ([8], see Figure 3). The robot uses speech synthesis and a laser pointer to present the multimedia content on a computer screen. For our own virtual presenter we also developed a own mark-up and synchronization language to generate multimodal expressions with separate channels for

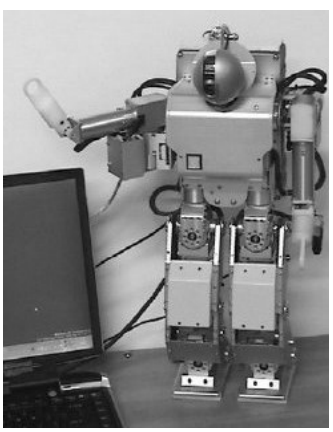

Figure 3. Robot presenter

deictic gestures, verbal expressions and sheet control (see [6,14] for more details). More interesting of course is to be able to generate a script from the contents of a presentation, e.g., the text on power point sheets. In the Jabberwocky system [3] keywords and phrases are extracted from the slides and during the presentation the system listens to the speaker and tries to switch slides at appropriate moments.

For our presenter (see Figures 4 and 5) a presentation script language has been designed. This MultimodalSync language represents a presentation on the level of synchronized multi-modal expressions. Example animation scripts have been obtained from annotated presentations taken from the corpus of meetings. The language has different layers that are executed in parallel, where synchronization is obtained through timestamps in the different layers. Layers that have been defined for power point presentations are the verbal layer (text on sheets), the sheet control layer and the deictic layer (sheets contain pointable areas). The projector screen that is used by the virtual presenter is a visual 3D entity in a virtual environment (meeting room) displaying sheets.

PUTTING VIRTUAL PRESENTERS IN VIRTUAL ENVIRONMENTS

In Figure 5 we already saw our virtual presenter in a virtual environment that allows presentations on a screen. Pictures, photographs, paintings, and sequences of these objects (e.g., a power point presentation) can be displayed on this virtual screen, and the virtual presenter can be provided with a script to explain and give comments. However, where is the audience? Clearly, we can introduce a fully generated virtual audience, as has been done in the experiments of Slater [9]. However, it is much more interesting to make a link with a real audience. We make a distinction between three types of audiences:

- An audience that is physically present during a presentation that is done by a human presenter in the same room;
- An audience that is provided with a presentation done by a virtual presenter without the ability to interact or to become aware of others that are accessing the same presentation;
- An audience that is physically present during a presentation that is done by a human presenter in the same room but that has also real-time access to a virtual reality representation of what is going on during the presentation, maybe allowing them to get additional information about the presentation;
- An audience that in addition consists of one or more persons that are not physically present in the presentation location, but that can take a virtual position in this audience from a remote location and from that location have a real-time view on things that are going on in the presentation location. They can be represented visually as virtual humans, observable for other members of the audience. In more advanced applications they also can take part in discussions and interactions with the virtual presenter;

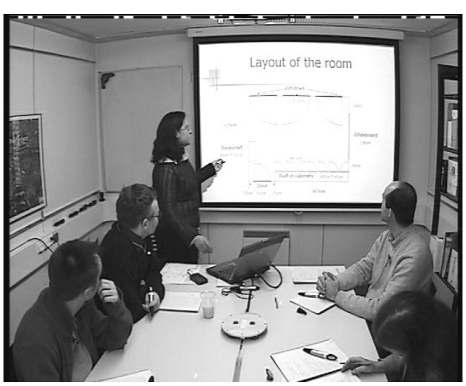

Figure 4. Meeting room presentation

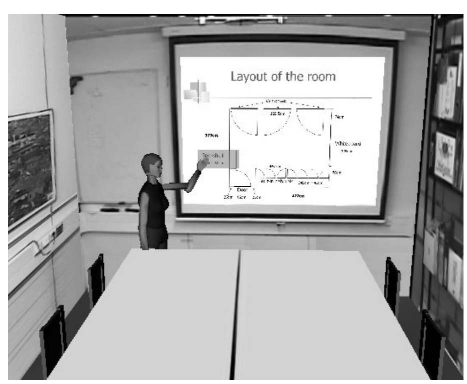

Figure 5. Virtual meeting room

- An audience that is interested in what has been going on during a presentation in the past. In this case we assume that the audience has facilities to browse through the presentation event, to retrieve information about the presentation event, and to ask meta-questions about the presentation event (who were present?, who asked a question about what?, was there concensus?, was there someone who didn't agree?, etc.).

In our research in the AMI project our first aim is to model a virtual presenter, to allow users access to presentations and, much more interesting, to put the presenter in a virtual environment allowing us to represent the presenter's audience in various ways (off-line representation of the physically present audience, real-time representation of the physically present audience, allowing remote participation of the event, etc.). In addition, we have looked at ways to translate the behavior of a human presenter and a human audience (assuming a presentation by a human presenter in a physical environment) to a virtual reality presentation event. That is, the movements of the audience members and a human presenter are captured by cameras and mapped on an H-Anim representation of the human body and the limbs that can be visualized as virtual humans in a 3D virtual environment. In our AMI project the results are visualized in a virtual meeting room. Once we are able to make this translation it is also possible to have different viewpoints on a presentation event or to let an observer navigate in the environment. A remote participant, whether represented and made visible to other participants or not, can take a reserved position during the event and an off-line user can experience a presentation event from the point of view of one of the real-time participants.

CONCLUSIONS AND FUTURE RESEARCH

In this paper we have looked at research on presentation modeling. Although the research takes place in the context of a European project on meeting modeling, the results (models and tools) can be used in other domains as well. We can learn from real presenters. That is why we aim at introducing annotation tools and schemes to annotate real presentations and that is why we design models that describe useful human presentation behavior. Another way to make use of real presentations is to have them real-time transformed in real-time to a presentation by a virtual presenter in a virtual environment that can be accessed by everyone who is invited or otherwise interested. Yet another way is to use motion capturing in order to obtain smooth animations of gestures, movements and pointing behavior of a virtual presenter. Presently we take two roads: (1) real-time translation of the behavior of a human presenter into the behavior of a virtual presenter in a virtual environment and (2) have scripted presentations performed by a virtual presenter without having a human counterpart [6,14]. One important line of future research is to design virtual presenters that can be interrupted and are able to handle simple interruptions. Obviously, there are multiple possibilities of interruption (requesting additional information, requesting the rephrasing of information, going back or forward to another position in a power point presentation) and the presenter has many possibilities to react (elaborating on a specific element of the presentation, giving an alternative presentation of a specific topic, move forward or backward in a presentation, informing the audience that the requested information will be given later or that it is not available, requesting additional information from the audience). A model of the process of interrupting a presenter, based on a taxonomy of such interruptions, is in development. The model that is in development describes how the different interruptions lead to their respective responses [13].

Other topics of future research are realistic gaze behavior of the virtual presenter, in particular when the presenter is interrupted. It certainly is not difficult for the reader to come up with a lot of other interaction issues that need to be modeled in order to obtain more realistic virtual presenters. Fortunately, it is not always necessary to simulate all human characteristics in order to obtain an embodied agent that is sufficiently believable for its human partners.

References

[1] AMI: Augmented Multi-party Interaction. http://www.amiproject.org/, 2004.

[2] S. Descamps & M. Ishizuka. Bringing affective behavior to presentation agents. Proc. 3rd Int'l Workshop on Multimedia Network Systems (MNS2001) (IEEE Computer Soc.), Mesa, Arizona, 2001, 332-336.

[3] D. Franklin, S. Bradshaw & K. Hammond. Jabberwocky: You don't have to be a rocket scientist to change slides for a hydrogen combustion lecture. In Proceedings of Intelligent User Interfaces 2000 (IUI-2000).

[4] N. Magnenat-Thalmann & P. Kalra. The Simulation of a Virtual TV Presenter. Proc. *Pacific Graphics 95*, World Scientific, Singapore, 1995, 9-21.

[5] A. Nijholt. Smart Exposition Rooms: The Ambient Intelligence View. In: Proceedings *Electronic Imaging & the Visual Arts (EVA 2004)*, V. Cappellini & J. Hemsley (eds.), Pitagora Editrice Bologna, March 2004, 100-105.

[6] A. Nijholt, H. van Welbergen & J. Zwiers. Introducing an embodied virtual presenter agent in a virtual meeting room. In: Proceedings IASTED International Conference on Artificial Intelligence and Applications (AIA 2005), Innsbruck, Austria, February 2005.

[7] T. Noma and N.I. Badler. A virtual human presenter. Proc. *IJCAI-97 Workshop on Animated Interface Agents*, 1997, 45-51.

[8] Y. Nozawa, H. Dohi, H. Iba & M. Ishizuka. Humanoid Robot Presentation Controlled by Multimodal Presentation Markup Language MPML. Proc. 13th IEEE Int'l Workshop on Robot and Human Interactive Communication (RO-MAN2004), Kurashiki, Japan, 2004.

[9] D.-P. Pertaub, M. Slater & C. Barker. An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators and Virtual Environments* 11 (1), 68-78.

[10] T. Rist, E. Andre & J. Muller. Adding Animated Presentation Agents to the Interface. In: *Intelligent User Interfaces*, 1997, 79-86.

[11] J. Terken (ed.). Report on Observation Studies with Requirements for CHIL Services. Deliverable D7.1 of the Project CHIL (Computers in the Human Interaction Loop) IP 506909, 14 July 2004.

[12] M. Theune, D. Heylen & A. Nijholt. Generating embodied information presentations. Chapter in: *Intelligent Information Presentation.* O. Stock & M. Zancanaro (eds.), Kluwer Academic Publishers, 2005, 47-70.

[13] J. Vlasveld. A feedback aware virtual human presenter. Internal HMI Report, University of Twente, December 2004.

[14] H. van Welbergen et al. A virtual human presenter in a 3D meeting room. Manuscript in preparation, University of Twente, March 2005.