Virtual Meeting Rooms: From Observation to Simulation

Dennis Reidsma, Rieks op den Akker, Rutger Rienks, Ronald Poppe, Anton Nijholt, Dirk Heylen, and Job Zwiers

Human Media Interaction Group
Center of Telematics and Information Technology
University of Twente, Enschede 7500 AE, The Netherlands,
dennisr@ewi.utwente.nl,
WWW home page: http://hmi.ewi.utwente.nl/

Abstract. Much working time is spent in meetings and as a consequence meetings have become the subject of multidisciplinary research. Virtual Meeting Rooms (VMRs) are 3D virtual replicas of meeting rooms, where the various modalities such as speech, gaze, distance, gestures and facial expressions can be controlled. This allows VMRs to be used to improve remote meeting participation, to visualize multimedia data and as an instrument for research into social interaction in meetings. This paper describes how these three uses can be realized in the various stages in the development of a VMR. We describe the process from observation through annotation to simulation and a model that describes the relations between the annotated features of verbal and nonverbal conversational behavior. As an example of using a VMR for research into how a single modality can influence the perception of social interaction, we present an experiment in the VMR where humans had to identify the speaker in a multi-party conversation in the VMR, based only on head orientations.

1 Introduction

Much working time is spent in meetings and as a consequence meetings have become the subject of multidisciplinary research. The introduction of technology in meetings offers new perspectives on, amongst others, communication and language, human perception and social interaction. In this paper we describe how virtual meeting rooms can be used to improve remote meeting participation, to visualize multimedia data and as an instrument for research into social interaction in meetings.

AMI is a European Research Project that aims at developing new technologies for supporting meeting activities such as meeting browsers, and technology that makes remote meeting participation easier, more effective and more natural.[1] The Human Media Interaction (HMI) group of the University of Twente is

[1] AMI - Augmented Multi-party Interaction.

one of the AMI partners [Nijholt et al., 2004]. The HMI group has a tradition in research into interaction with embodied conversational agents, computer graphics for virtual environments and machine learning techniques for recognition of higher level features (e.g. dialogue acts, gestures and emotions) from lower level features (e.g. words, hand movements and facial features).

The paper is organized as follows. In Section 2 we will present our view on the development of the concept of meeting in interaction with technological developments. In Section 3 we present a schematic overview of the process from observation to simulation and we discuss possible uses of virtual meeting rooms in Section 4. As an illustration of the scheme we focus in Section 5 on some experiments we did with a virtual meeting room.

2 Meetings in virtual reality

In a general sense a meeting is any coming together, willingly or unwillingly, of two or more people at such a close distance of each other that they are aware of each others presence and, willingly or unwillingly, react on that. The concept of distance, and related to that the concept of being in the same meeting room, has strongly been developed and is still being renewed by the development of technology in the last few centuries, in particular by developments in communication and information technology. This is really a process of conceptual development, in which the content of sharing the same space evolves from physically sharing the same space to mentally sharing the same space. We identify invariantly a number of central themes: the struggle for the individual privacy, respecting each others private space, the need of being respected by others, the will to express one self and one's ideas and to realize individual goals. *In a more restricted sense* a meeting is an organized process of people coming together focussing on a common topic or task. Meeting in this sense is one of the characteristics of the modern way we organize our work in all kinds of organizations. However professionalized and organized a meeting may be, it is still a gathering of people. All the themes that play in the more general sense of meeting can be identified in these meetings as well, be it often in more organized, conventional forms, and mediated by invented rules of good conduct: turn taking behavior, addressing behavior, politeness rules, and dominance relations.

The impact of technology on meetings can not be described adequately in terms of quantitative measurable effects it has on properties of processes that occur in existing forms of meetings. Technology develops the very idea of meeting itself, and it has impact on how people realize the idea of meeting. Moreover, what is essential for meetings is that technology offers new perspectives on, amongst others, communication and language, human perception and social interaction. These new perspectives may help to gain more insight in the essential qualities of these aspects of social reality.

State of the art in computer graphics and embodied conversational agents allows the creation of *Virtual Meeting Rooms* (VMRs), virtual replicas of real meeting rooms. VMRs are useful for various purposes that can be grouped into the following three categories:

1. As an (immersive) virtual environment, a communication means for real-time remote meeting participation. Real-time presentation of a virtual model of a meeting reduces the amount of data that has to be sent to and displayed on client side remote displays. It will be clear that this builds on knowledge about what types of events and behaviors in the real meeting are essential to be presented in the virtual meeting in order to maximize the quality of those impressions that are required by the user given his task and role in the meeting, such as the feeling of presence[2], and the possibility of mutual gaze.
2. Presentation of multimedia information about meetings. Information can be directly obtained from recordings of behaviors in real meetings (e.g. tracking of head or body movements, voice), from annotations or from machine learning models that induce higher level features from recordings. One could also think of a 3D summary of real meetings. These presentations can be used for evaluation of annotations and results obtained by machine learning methods.
3. Research into human social interaction, recognition, and interpretation of visualized information. Virtual Environments allow control of various independent factors (voice, gaze, distance, gestures, facial expressions) and can be used to study how they influence features of social interaction and social behavior. Conversely, the effect of social interaction on these factors can be studied adequately in Virtual Environments as well.

In the remainder of this paper we describe how each of these purposes can be accomplished in a VMR. We constructed a 3D virtual replica of one of the meeting rooms that is used for data collection of meetings in the AMI project, based on our agent platform (implemented in *Java*, using *Java3D*, *X3D* and *XML*-technology, using *H-ANIM* standards for human body animation). In Section 5 we will demonstrate the results of an experiment that was carried out in this HMI VMR.

3 From observation to simulation

In this section we describe the process from observation through annotation to simulation and a model that describes the relations between the annotated features of verbal and nonverbal conversational behavior.

[2] "Presence (..) describes the cognitive process of constructing an environment. As a result of this construction, the user experiences a sense of presence, that is the user feels him- or herself as part of the virtual environment. Since the body is real, the 'realness' of the virtual environment is inferred. Users describe that they are 'there' and that the virtual stimuli can have actual effects on behavior and emotions." [Regenbrecht and Schubert, 1997].

Figure 1 shows an abstract view of the Virtual Meeting Room 'observation to simulation' process. The left hand side depicts the observation and interpretation. Human interactions in meetings are recorded on video and audio. Observation of these videos leads to descriptions of observable events (body movements, facial expressions, speech, etc). These observations can be interpreted on progressively more complicated levels (see also [Reidsma et al., 2005]). The right hand side depicts the simulation process. At a certain point, the information from the annotations is used for *(re)generation* of the communication behavior, recreating the lower level information from models of human interaction (see also Section 4.2).

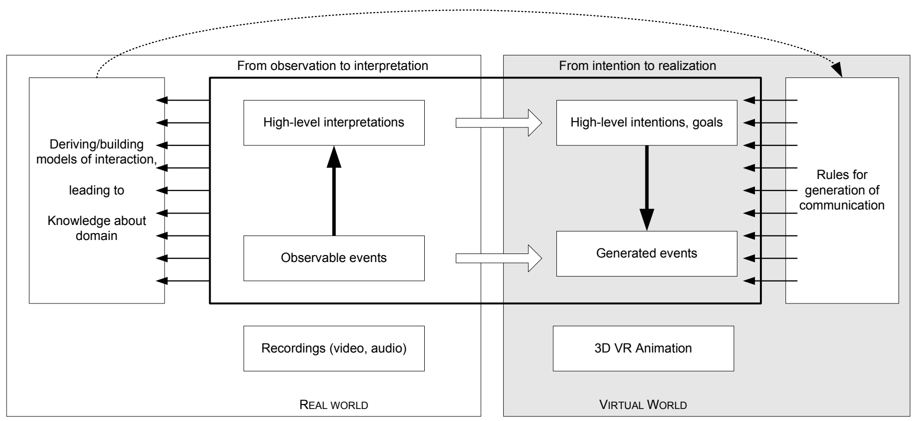


Fig. 1. Schematic overview of the various steps from observations and recordings via annotations to simulations mediated by various models expressing the relations between the aspects of verbal and non-verbal conversational behavior of participants in the meeting.

3.1 Annotations of behavior in meetings

The first step in realizing the process described in the previous section is annotation of recordings containing human-human interactions. Within the AMI meeting project we see a huge effort in meeting data collection, meeting data annotation and dissemination of these data for various multidisciplinary research purposes inside and outside the project. One hundred hours of meeting recordings are planned for of which about 60% are scenario based meetings with four people meeting four times. This is part of a design project in which they have to work on a prescribed task to develop a remote TV control unit. Participants have various roles in this play and in order to meet reality as best as possible, external events and information are brought in that may influence the decision making process as well as the outcome of the meetings.

The hundred hours of recordings will be annotated in varying levels of detail for different dimensions. There are several reasons for creating manual annotations of corpus material. In the first place ground truth knowledge is needed in order to evaluate new techniques for automatic recognition of those same aspects. In the second place, as long as the quality of the automatic recognition results is not high enough, only manual annotations provide the quality of information needed to do research on human interaction patterns (see also Section 4.3).

A few examples of layers that can be annotated are hand and body postures, labelled gestures (interpretation of movement and pose), speech transcription, communicative acts, argument structures / topics and summaries.

The annotations can be organized in layers of increasing complexity. The lowest layers describe mostly the *form* of the interactions, or the observable events. The higher layers describe interpretations of these observable events, giving the *function* of the interactions. Consider for example the situation where a participant raises his or her hand. The form of this gesture can be observed and annotated as 'hand raising'. On an interpretation layer, this event may be annotated with the function of this gesture, such as 'request for a dialogue turn' or 'vote in a voting situation'.

Once the annotations are available they will be analyzed. One of the results of such analyses will consist of models of human interaction on varying levels of abstraction. Lower level models might describe how people generally realize certain communicative goals (e.g. how to express the addressee of utterances, or how to show disagreement or agreement). Higher level models might describe aspects such as what interaction patterns characterize efficient meetings.

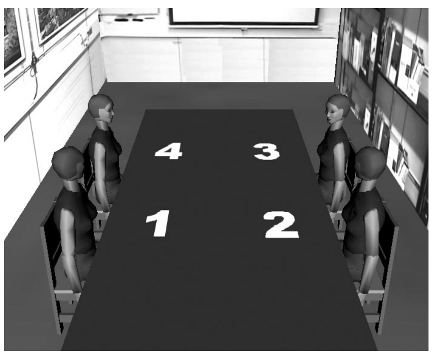
3.2 Regeneration of behavior in meetings

The annotations, together with selected models derived from these annotations, can be used to replay (parts of) a meeting in a *Virtual Meeting Room*. Figure 2 shows an image of the real meeting room together with three different views of the HMI VMR. The annotations described in the previous section can be replayed in the meeting room in different ways. Replay can show all available annotated information (down right in picture, a shot that shows head orientation, recognized body pose, current speaker and addressees of utterance) or only a selection (down left shows for example only head orientation).

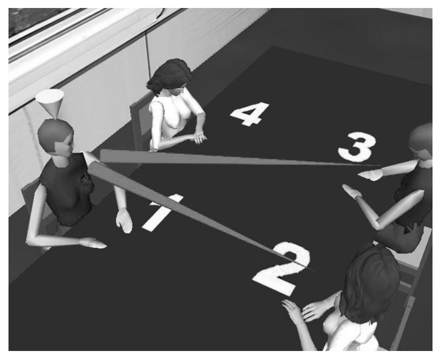
Furthermore the replay can be a direct replay of observed behavior, or an interpreted replay, starting from high level interpretation of what happened during the meeting, generating appropriate behavior that expresses the right content but in a potentially different form. The rules for generation of communication are derived from domain knowledge (models and theories of human interaction) collected through the analysis of large amounts of data from real world examples. Examples are models for choosing modalities, realizing gestures or speech, formulating sentences, deciding on communicative goals given beliefs and intentions, choosing communicative actions based on goals, etc.

(a) Real setting of meeting 3, with participants with flock of birds sensor for recording exact head movements

(b) VMR view from the eyes of a participant



(c) VMR central view



(d) VMR extended view with visualization of head orientation, body pose, speaker and addressees

Fig. 2. Real and virtual meeting room

4 Uses of the Virtual Meeting Room

This section goes into more detail concerning some of the uses of the VMR. In Section 2 it was already mentioned that this paper will focus on three categories of VMR applications: an environment for teleconferencing that provides a sense of immersion and presence, visualization of multimedia information from meetings for several purposes and an instrument for elicitation and validation of models for social interaction.

4.1 Remote participation and enhancement of meetings

A Virtual Meeting Room can be used as an environment for teleconferencing [Greenhalgh and Benford, 1995]. In addition to the usual advantages of remote meeting participation it offers control over some features that are problematic in traditional video-based conferencing (e.g. natural visualization of gaze direction cues). But there are more opportunities for influencing the remote interaction in a teleconference in a Virtual Meeting Room.

In the first place different meeting participants need not necessarily have the same view of the Virtual Environment. This simple fact introduces a lot of possibilities worth investigating. Participants can adapt the Virtual Environment in which the meeting takes place to their own preferences and comforts without disturbing the other people. Each person can be given his or her own perception of the seating arrangements. Since it is known that some positions are more advantageous in terms of discussion impact than others, it might be sensible to give each participant such a view of the seating that he or she never feels to be in the most disadvantageous position, leading to all participants feeling more comfortable during the meeting. Another way of adapting the meeting to ones own preferences involves Transformed Social Interaction [Bailenson et al., 2004] which allows a participant to influence the way that he or she is presented remotely.

A Virtual Teleconferencing Environment also offers the possibility to introduce autonomous agents that have the same communicative channels at their disposal as the human participants. This gives opportunities for defining experiments to discover regularities in human social interaction, as will be described in Section 4.3. It also facilitates the introduction of *helper agents* into an actual meeting. Existing work has already shown that people can be influenced in their behavior as well as their assessment of a situation by the presence of autonomous agents and their behavior, even if they know that the agents are not representing a real human [Pertaub et al., 2002]. The emergence of advanced recognition technology for human interaction, partly developed from extensively annotated corpora, will allow Embodied Conversational Agents (ECAs) to use this fact to influence the course of the meeting. A simple example would be the introduction of a virtual chairman in the meeting room with a regulating task. Based on an analysis of what is going on in the meeting, the virtual chairman can influence the progress of the meeting (request a vote, encourage silent people to speak, mention gaps in the argumentation, etc). An enhancement of this chairman would be possible if the recognition technology gets advanced enough to detect potentially tense situations: the virtual chairman could try to defuse such situations by making a joke, or changing the subject of discussion. Another example is to include a virtual participant who listens very attentive whenever person A is saying something, and gets bored and restless whenever person B is saying something, in order to increase the status and believability of person A.

4.2 Re-visualization of meetings

Using a general implementation of a VMR it is possible to re-visualize the contents of a recorded meeting. This can be done literally, trying to stay as close to the original recordings as possible, or more conceptually, aiming for a visualization that reflects the meaning of the meeting rather than the actual form.

The re-visualization process traces a path through Figure 1. This path starts at the bottom left corner (real world / video recordings), and first goes upwards through various stages of observation and interpretation. At a certain point the transition to the right part of the model is made (in a sense 'copying' the information present on one level from the left hand side to the same level on the right hand side), after which the generation flow is followed down to produce an animation of the meeting in the virtual meeting room (bottom right).

Transition at the lowest levels (Figure 3) is already interesting: replaying recognized 3D joint angles in a VMR in parallel with showing the original video offers a kind of quick-and-ready validation of the pose recognition process. If the recognition is good enough to use as input for a gesture labelling algorithm but not good enough to give convincing replay results, the transition can be made at a higher level. After interpreting the movements as labelled gestures, the replay is created from these gesture types rather than directly from the body poses, leading to an animation that is not an exact replica of the original video but does express the *meaning* of the movements possibly more clearly.

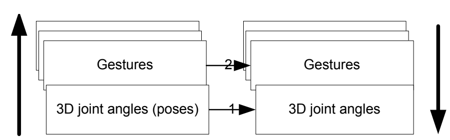


Fig. 3. From observation to simulation on different levels of interpretation

Another possible level where the transitions can be made is the level of communicative actions such as contributions to or judgements about the current topic of discussion. The simulation on this level might be able to use different realizations for the same communicative actions. This can be useful if one wants to use the appropriate culturally determined gestures, or to highlight aspects of the contributions in relation to social conventions.

The final and a much more complex possibility discussed here deals with *summarized replay* of a meeting or set of meetings. If a discussion about a certain issue is spread over fragments of several meetings, at a certain level of interpretation the main structure of the arguments can be found. Making the transition at this level, selective replay enables a new *cohesive* and *interpreted* replay of the discussions. If the models for simulating interaction are good the main points of the original meeting will stay intact (who proposed what, who was for / against, who used / supported which arguments, etc), without the redundant information that was conveyed in reality. This form of simulation will deviate much from the original recordings, but the relevant content (the function) remains the same.

4.3 Validations of models of social interaction

This section describes some of the applications of the VMR most related to social intelligence. If autonomous agents must display believable social behavior, there are many communicative aspects to be taken care of. For such aspects models are needed: in what circumstances are which communicative actions desirable? How does a person show whom he / she is addressing? Does it depend on status differences? What is acceptable behavior for an ECA to show that he / she is listening to the user and interested in what the user says? How do people exhibit and perceive signals related to relative status? The same type of models will also be needed for effective automatic analysis of meetings for other purposes such as retrieval or meeting support. The VMR provides ways to both elicit and validate such models. The following paragraphs give a few examples of this, of which at least the addressing experiment will actually be done in the HMI VMR. A few other experiments that use Virtual Environments for elicitation and/or validation of models of (social) interaction can be found in Bailenson et al. [2001] and Slater and Steed [2001].

VMR Turing test The VMR Turing test (adapted from Bailenson et al. [2004] allows one to validate a complex set of models, testing whether they result in convincing, natural social interaction by ECAs.

It works as follows: show a human subject a VMR with avatars controlled by other humans and avatars controlled by an ECA. Remove from the human avatars all communication channels that the ECA does not have (for example face expressions). Ask the subject to judge which avatars are controlled by the ECA. This can be done for any level of complexity (for example, you can validate models of listening behavior by having the subject talk to two humanoids of which one is ECA and one is human and see whether the subject can tell which is which, if both are not allowed to talk back).

Validate models of addressing behavior Jovanović and op den Akker [2004] give an outline of their plans for research on modelling and detection of addressee in conversation. When such models have been developed, based on corpus annotations, these will be validated in the HMI VMR. A possible way to achieve this is to let an ECA simulate a fragment of conversation, expressing the addressee of utterances in one of the many ways allowed by the model (using vocatives, gaze, etc.). A human participant, immersed in the VMR, will then be asked to assess who is the addressee of utterance. This experiment can provide the validation whether a model of addressing behavior is good enough to use in an ECA, insofar as that a human will understand its addressing cues.

Turn taking A experimental setup similar to the one described for validation of addressee models can be used to validate certain patterns of nonverbal behavior related to *turn taking*, simulated in Padilha and Carletta [2003]. The simulation models can be applied to generate turn taking behavior in a number of ECAs. A human participant can be instructed to assess which person is the next speaker, the previous speaker, etc. Models that lead to a better prediction are more suitable for use in ECAs.

5 Example of an experiment in the VMR: speaker prediction from head orientation

Social intelligence is very much related to both the understanding and generation of nonverbal communication. Our VMR allows us to analyze nonverbal communication and its social intelligence aspects between the inhabitants of a VMR, and it allows us to generate and validate social interaction behavior from models of social intelligence. Since different communication channels (gestures, head movements, facial expressions, etc.) can be controlled in the VMR, we are able to zoom in on the social intelligence properties of one particular modality, to leave out other modalities, and to study any combination of modalities [Loomis et al., 1999]. We are particularly interested in generating believable models of speaker and addressee behavior in meetings. To gain more insight into this behavior, we are studying the impact of various modalities on the prediction of both speaker and addressee. Here we describe an experiment where humans were asked to predict the current speaker, solely given the head orientations of all participants at a given time in a meeting.

A number of researchers reported studies concerning the functions of gaze and mutual gaze in conversations. According to Kendon [1967] gaze serves four functions: visual feedback, regulate conversational flow, communicate emotions and relationships and to improve concentration by restricting visual input. Gaze behavior of speakers, addressees and overhearers is related to turn-taking and turn-giving behavior as well as to addressing behaviors, behaviors that speakers show when they are addressing their speech to one or a selected subgroup of participants in a meeting (see e.g. Jovanović and op den Akker [2004]). Gaze is defined as the direction where the eyes are pointing in space. It is the sum of head orientation and eye orientation [Gibson and Pick, 1963, Stiefelhagen and Zhu, 2002]. Since recording eye gaze without being obtrusive is hard (see e.g. Vertegaal [1998]) *head orientation* alone is often used as indication of gaze and focus of attention. In a studies with four meeting participants at a round table, Stiefelhagen [2002] reported that in 89% of the time it could be correctly determined at whom the subject was looking based only on the participant's head orientation.

In the smart meeting room at IDIAP in Martigny three four-person group meetings have been recorded where participants had the task of debating several statements[3]. Participants wore an electro-magnetic sensor on their heads so that the exact head position and orientation could be recorded. The meetings were further audio and video recorded. After recording, biases in head orientation due to incorrect mounting of the sensor on the head were removed. From the

[3] These meetings do not belong to the *hub* AMI meeting data collection, they are *spoke* recordings, for research interests of individual AMI project partners.

obtained data set, all occurrences with non-speech (laughter, silence, etc.) or with speech overlap were removed. Table 1 shows the number of frames for each meeting, the prior speaker probabilities of $P(Speaker = person_i)$ for each of the 3 meetings separately and for all meetings in total.

	Meeting 1	Meeting 2	Meeting 3	Total
Samples	11333	13078	28148	52559
A priori person 1	40.4%	26.9%	24.8%	28.7%
A priori person 2	27.3%	23.4%	9.8%	16.9%
A priori person 3	7.7%	8.6%	29.4%	19.5%
A priori person 4	24.7%	41.2%	36.0%	34.9%

Table 1. Number of samples and *a priori* speaker probabilities

5.1 Speaker prediction with a Naive Bayes classifier

To predict the speaker solely using head orientations one could use the maximum *a posteriori* probability of a person speaking given the head orientation of all participants. To determine which person was being looked at by another person we used boundaries in the azimuth angle range.

We trained a Naive Bayes classifier with supervised discretization [Dougherty et al., 1995] to determine the optimal decision boundaries. Then a test was conducted for three different meetings using tenfold cross validation. It appeared that within a single meeting the classifier performed quite well (82.9%). However when our training and test sets were taken from different meetings the performance dropped significantly (to 35.1%). For a more elaborate explanation and discussion the reader is referred to Rienks et al. [2005].

5.2 Speaker prediction by humans

A second experiment was carried out where humans had to perform the same task. They had to decide who was the speaker when presented a view of the VMR. Participants of the experiment were shown the meeting room with the participants, displaying only azimuth head angles (see Figure 2.). There was an option panel where they were able to choose amongst the four speakers. To prevent participants from conveying indifference to the task a 'no idea' button was available [Ray, 1990].

Each experiment consisted of a session with four parts, each containing 20 samples. There were two types of sessions, with and without feedback. The feedback was only given on the first session part by showing a red arrow above the head of the correct speaker. The first two parts of both session types were sampled randomly from meeting 3, the third part contained 20 randomly chosen samples from meeting 2 and the last part contained 20 randomly chosen samples from meeting 1.

The idea behind performing the experiment with and without feedback was twofold. In the first place it enabled us to see if the feedback was helpful to the participants. Secondly, we were able to see whether feedback on samples from one meeting influenced the results on samples from other meetings. The meeting participants were students and employees of our department. The two session types were each completed 20 times, resulting in a total of 3200 answered samples. The results are shown in Table 2.

	Part 1	Part 2	Part 3	Part 4	Total
With feedback	47.8%	49.3%	29.8%	24.8%	37.9%
Without feedback	39.3%	42.0%	33.3%	35.3%	37.4%

Table 2. Classification results for humans per session type

The table shows that the human performance was approximately 38%, which is lower than we expected. An interesting finding here is the significant difference between the results on the first two session parts. The first two session parts were better answered with feedback than without feedback ($t(38) = 2.29; p < 0.05$) and for the last two session parts we see the opposite ($t(38) = -2.95; p < 0.01$). Furthermore it appeared that when no feedback was given the performance remained more stable over the different session parts. When no feedback was given, the experiment participants were not informed of the *a priori* speaker probabilities. The participants that received feedback on the first session part might have 'learned' the *a priori* distributions of meeting 3 by adapting to the feedback. For a more elaborate explanation and discussion the reader is again referred to Rienks et al. [2005].

6 Conclusions and Further research

Virtual meeting rooms may add value to the already existing technological means people have to communicate and meet. The various modalities such as speech, gaze, distance, gestures and facial expressions can be controlled, which allows VMRs to be used to improve remote meeting participation, to visualize multimedia data and as an instrument for research into social interaction in meetings. We described the process from observation through annotation to simulation and a model that describes the relations between the annotated features of verbal and nonverbal conversational behavior. This model can be used to relate various research tasks in the field of meeting research. An experiment was conducted in the VMR where humans had to predict to speaker based on head orientations. It appeared that there are significant differences in gaze behavior of speakers and listeners when looking only at the modality of head orientations.

Several questions remain concerning research on human performance in recognizing the speaker and more general the flow of conversation. Comparisons with experiments using real pictures of the meeting instead of virtual reality scenes may prove interesting. We will also perform similar experiments using another selection of channels (e.g. hand and arm movements or body poses). Further, we will perform similar experiments to see how good humans are in deciding who is the *addressee* in a given situation showing head movements (with or without real postures and gestures). We will compare these results with machine learning techniques trained on annotated meetings. Then we will pursue our work on meeting modelling and see how we can present real meetings in an effective way by means of a virtual representation that shows the most informative view on the meeting.

A lot of research remains to be done to see how people perceive and interpret meeting situations and how they react on them in a virtual meeting room. Results of such research are necessary to see what information channels and modalities are important to effectively perform the various tasks in a meeting. This concerns not only the transfer of task-based information, but also issues such as maintaining a good feeling of social presence by representing the appropriate communicative cues.

Acknowledgements This work was partly supported by the European Union 6th FWP IST Integrated Project AMI (Augmented Multi-party Interaction, FP6-506811, publication AMI-42).

Bibliography

J.N. Bailenson, J.J. Blascovic, A.C. Beall, and J.M. Loomis. Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence: Teleoperators and Virtual Environments*, 10(6):583–598, 2001.

J.N. Bailenson, A.C. Beall, J. Loomis, J.J. Blascovich, and M. Turk. Transformed social interaction: Decoupling representation from behavior and form in collaborative virtual environments. *Presence: Teleoperators and Virtual Environments*, 13(4):428–441, August 2004.

J. Dougherty, R. Kohavi, and M. Sahami. Supervised and unsupervised discretization of continuous features. In A. Prieditis and S. Russell, editors, *Proceedings of the 12th International Conference on Machine Learning*, pages 194–202, San Francisco, CA, 1995. Morgan Kaufmann.

J.J. Gibson and A.D. Pick. Perception of another persons looking behavior. *American Journal of Psychology*, 76(3):386–394, September 1963.

C.M. Greenhalgh and S.D. Benford. Virtual reality tele-conferencing: Implementation and experience. In *Proceedings of the Fourth European Conference on Computer Supported Cooperative Work (ECSCW95)*, pages 163–178, September 1995.

N. Jovanović and R. op den Akker. Towards automatic addressee identification in multi-party dialogues. In M. Strube and C. Sidner, editors, *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue*, pages 89–92, Cambridge, MA, April 2004. Association for Computational Linguistics.

A. Kendon. Some functions of gaze direction in social interaction. *Acta Psychologica*, 26(1):22–63, 1967.

J.M. Loomis, J.J. Blascovich, and A.C. Beall. Immersive virtual environment technology as a basic research tool in psychology. *Behavior Research Methods, Instruments and Computers*, 31(4):557–564, November 1999.

A. Nijholt, R. op den Akker, and D. Heylen. Meetings and meeting modeling in smart surroundings. In A. Nijholt and T. Nishida, editors, *Social Intelligence Design. Proceedings third international workshop*, CTIT Workshop Proceeding Series WP04-02, pages 145–158, Enschede, The Netherlands, July 2004. ISBN 90-75296-12-6.

E.G. Padilha and J. Carletta. Nonverbal behaviours improving a simulation of small group discussion. In *Proceedings of the First International Nordic Symposium of Multi-modal Communication*, pages 93–105, Copenhagen, Denmark, September 2003.

D.P. Pertaub, M. Slater, and C. Barker. An experiment on public speaking in response to three different types of virtual audience. *Presence: Teleoperators and Virtual Environments*, 11(1):68–78, February 2002.

J.J. Ray. Acquiescence and problems with forced-choice scales. *Journal of Social Psychology*, 130(3):397–399, 1990.

H.T. Regenbrecht and T.W. Schubert. Measuring presence in virtual environments. In *Proceedings of the Human Computer Interaction International Conference*, San Francisco, CA, 1997.

D. Reidsma, R. Rienks, and N. Jovanović. Meeting modelling in the context of multimodal research. In *Proceedings of the MLMI'04 workshop*, volume 3361 of *LNCS*, pages 22–35, Martigny, September 2005. Springer Verlag. ISBN 3-540-24509-X. ISSN 0302-9743.

R.J. Rienks, R.W. Poppe, and M. Poel. Speaker prediction based on head orientations. In M. van Otterlo, M. Poel, and A. Nijholt, editors, *Proceedings of the Fourteenth Annual Machine Learning Conference of Belgium and the Netherlands (Benelearn 2005)*, CTIT Workshop Proceedings Series WP05-03, pages 73–79, Enschede, The Netherlands, February 2005. ISSN 1574-0846.

M. Slater and A. Steed. *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*, chapter Meeting People Virtually: Experiments in Shared Virtual Environments, pages 146–171. Springer-Verlag, London, UK, 2001. ISBN 1-85233-461-4.

R. Stiefelhagen. Tracking focus of attention in meetings. In *Proceedings of the IEEE International Conference on Multimodal Interfaces*, pages 273–280, Pittsburgh, PA, October 2002.

R. Stiefelhagen and J. Zhu. Head orientation and gaze direction in meetings. In *CHI'02 extended abstracts on Human factors in computing systems*, pages 858–859, Minneapolis, MN, April 2002. ACM Press. ISBN 1-58113-454-1.

R. Vertegaal. *Look who's talking to whom*. PhD thesis, University of Twente, Enschede, The Netherlands, September 1998.