

# Meetings in the Virtuality Continuum: Send Your Avatar

Anton Nijholt

*Center of Telematics and Information Technology,  
University of Twente, Enschede, the Netherlands.  
anijholt@cs.utwente.nl*

## Abstract

*We survey our research on verbal and nonverbal interactions between meeting participants. The participants meet in a smart environment where their activities are captured and interpreted by the environment. In this way the environment not only allows off-line access and retrieval of meeting information, but also real-time support of meeting activities. Since in our research our aim is to have real-time transformation of activities in the physical meeting environment to a virtual meeting environment, we also design our environment to allow real-time participation by remote meeting participants and by intelligent virtual meeting assistants that can support the meeting participants.*

## 1. Introduction

In this paper we survey our research on verbal and nonverbal interactions between meeting participants. The participants meet in a smart environment where their activities are captured and interpreted by the environment. In this way the environment not only allows off-line access and retrieval of meeting information, but also real-time support of meeting activities. Since in our research our aim is to have real-time transformation of activities in the physical meeting environment to a virtual meeting environment, we also design our environment to allow real-time participation by remote meeting participants (maybe represented as avatars) and by intelligent virtual assistants that support the human meeting participants.

In section 2 of this paper we discuss the AMI project. The emphasis is on the design of a corpus of meeting interactions and tools to annotate this corpus in order to analyze the data and to achieve models that underlie the data. Section 3 is devoted to our attempts to introduce off-line and on-line support to meeting activities. In order to provide support the activities in the smart meeting environment need to be interpreted and pro-active and reactive support will be based on access to this interpretation. A virtual reality representation of meeting activities en-

riched with useful meta-information is a very advanced form of interpretation. In this paper, in section 4, we discuss how we can obtain such a representation from captured data (using cameras and microphones) and we discuss also how it can play a useful role: for research purposes, for browsing meeting information and for remote participation, that is, a desktop, augmented reality or immersive virtual reality environment where participants meet and participate [4].

In particular we need to make clear how the environment can be inhabited by interacting human and virtual meeting participants. As a preliminary example of a virtual meeting assistant we discuss some work in progress on a virtual presenter. Some conclusions and future work are mentioned in section 5.

## 2. The Augmented Multi-party Interaction (AMI) project

AMI is a research project in the European 6th Framework program. AMI is concerned with research on multimodal interaction, and, as the name suggests, multimodal interaction in a multi-party context, a context where we have two or more persons interacting with each other and/or with smart entities (objects, virtual humans, robots, etc.) present in the environment. Obviously, this kind of research fits into paradigms of interactions in smart environments, ubiquitous computing, disappearing computers, and ambient intelligence. The AMI project concentrates on multi-party interaction during meetings. Obviously, there is multi-party interaction in educational settings, in class rooms, in offices, in workspaces, in home environments and in public spaces. It is assumed that models and technology developed in the meeting context will be useful in these other situations as well.

### 2.1. Development of tools and models in AMI

The main aims of the AMI project are to develop technologies for the disclosure of meeting content and to provide online support of (possibly remote) meetings. Partly these aims require the same technology. Intelligent

browsing and summarization of a previous meeting or previous meetings requires understanding of interactions and the content of interactions. Understanding interactions and the content of interactions is also necessary in order to provide real-time support to meeting participants. Remote participation requires means to take part in verbal and nonverbal interaction, that is, among others, means to mediate the turn-taking process and the gaze behavior of participants in order to allow smooth interactions. Hence, for off-line support, real-time support, and remote participation, we need models and model-based tools for multi-party interaction.

One of the issues we address in our research is the design of meeting assistants that can play a role during a meeting. They may help individual participants (did I make my points clear, what do others think of it), participants with a specific role (e.g., the chairperson) or all of the meeting participants (do we converge to a joint opinion or decision).

## 2.2. Corpus design and collection

In the AMI project meeting data is captured in smart meeting rooms at IDIAP (Martigny), at the University of Edinburgh, and at TNO Human Factors (Soesterberg). The meeting data in these meeting rooms is captured using different kinds of sensors: videos, microphones, whiteboard, PowerPoint presentations and smart pens. The emphasis in the project is on speech and image processing and how to fuse information coming from these two media sources in multi-media presentation tools. The main corpus consists of one hundred hours of video/audio registration of meeting activities, captured from some global video cameras, global microphone array arrangements, individual cameras and (lapel) microphones.

In order to collect data and to study meeting activities meetings have been arranged. Mock-up meetings collected in a previous project are available. Participants of these meetings were asked to follow a script that described the global theme and the global structure of their meeting, but there were no constraints on the way they gave content to their contribution. These meetings have been used in order to obtain and evaluate multi-party interaction models, e.g., models that try to capture ways meeting participants address their audience.

In the AMI project new and more natural meetings were recorded and rather than scripts, scenarios were used to have more natural interaction between the participants of these meetings. See Figure 1 for a global camera view of such a meeting. The scenarios that were used allowed design meetings, meetings in which the participants were asked to play different roles, e.g. a project manager, a marketing expert, a user interface designer and an industrial designer. How does such a development team agree about the design of a remote control?



Figure 1. Global camera view of meeting room

## 2.3. Corpus annotation and corpus annotation tools

How does a development team agree about the design of a remote control? In order to answer a question like this we need to study the information that is available in the corpus that has been collected. To study the corpus we need to design annotation schemes and this is an activity that requires interaction between observed data in the corpus and models that underlie the design of annotation schemes. Annotations apply descriptions to data. These descriptions can apply to phenomena at an individual level or at a group level. And, preferably these descriptions are based on theoretical models (of interaction) or they have been chosen because they are useful for the particular domain of application. Obviously, annotations as well as the theoretical models can describe meeting activity at different levels, for example, names of meeting participants, parts of speech, dialogue acts, gestures that are made, speaker, head and gaze orientation, who is addressed, focus of attention of a particular person or the group, displayed emotions, and current topic.

Tools, based on theoretical models, are available to obtain, although not with a hundred percent accuracy, some of these annotations automatically. This is especially true for annotations based on models from computational linguistics, dialogue modeling, low-level image processing and multimodal tracking. Tools for real-time support or off-line information access can be based on this automatically obtained information from a meeting and their sophistication and usefulness directly depends on the state of the art of automatically collecting this information and the intelligence needed to interpret this information. In order to advance this state of the art it is necessary to collect data and annotate the data manually. These annotations are useful for analysis of the data and the design and evaluation of more advanced theoretical models that describe the issues we are interested in and

that show in the data. Although those annotations have to be done manually, we can of course develop tools that allow efficient creation of annotations and tools in which knowledge about the phenomena to be annotated can be embedded. This embedding allows the tool to suggest annotations, to limit the choices of an annotator or to pre-fill values of attributes.

Among the annotation tools we developed are a Dialogue Act coder (DA) and a Continuous Video Labeling tool (CVL, [11]). These tools are based on the NITE XML Toolkit (NXT, [2]), a toolkit that allows display and analysis of cross-annotated multimodal corpora. In addition, NXT has a query language for the exploitation of annotated data. DA supports the annotation of dialogue acts, their addressees and relations between dialogue acts (adjacency pairs). Since dialogue act determination requires interpretation of the intentions of a speaker it is difficult to do this in real-time. DA supports the annotator by offering transcription text synchronized with video and by allowing browsing and replay of fragments. The CVL tool we developed supports real-time and off-line annotation of observations and interpretations from video. Examples are gaze direction, head orientation, postures and target of pointing gestures.

Generally, when looking at annotation creation, it is useful to distinguish between annotations that are based on observations (e.g., head nods) and annotations based on interpretation (e.g., dialogue acts and emotions). It is also useful to distinguish different layers of annotations, where a layer describes one particular way of annotating (e.g., gestures), and a layer can contain input for other layers e.g., a dialogue act layer can use input from a layer describing facial expressions or gaze. Labels for annotating can also be taken from an ontology describing the important concepts of the application. Segmentation of the input into fragments that can be referred to is another important issue.

Clearly, all these aspects have to be addressed when our research aims at making a change from manual annotation towards semi-automatical and fully automatical annotation and interpretation of data in order to provide real-time support to the inhabitants of a smart environment. Apart from designing models and rules from the observed and annotated events there is the possibility to train and learn the computer to build models that are able to automatically derive these annotations (that is, automatically find interesting and useful features for interpreting events).

## 2.4. Meeting modeling

A meeting model captures the knowledge that explains from a higher level the verbal and nonverbal multi-party interaction in the meeting context. There are individual and group activities that follow from the goals of a meet-

ing and from properties of being a group where individuals get together, have different roles, and may have different aims. These higher level phenomena show in relationships among the different features that are annotated and knowledge about them may determine the set of features to be annotated. Knowledge on the level of meeting models helps in predicting what is likely going to happen next in a meeting on a more global level.

Obviously, attempts can be made to structure meetings based on low-level phenomena. For example, in our project stochastic models (variants of Hidden Markov Models) are used to describe meetings as sequences of meeting actions. In such models too many important aspects of a meeting are not addressed. What does a meeting participant communicate or intend to communicate, when are there topic shifts and how can they be recognized, both from verbal and nonverbal cues and who is addressed by a speaker?

Group behavior and how individuals behave in a group setting is an important issue. This requires input from behavioral and social sciences [1,6]. Theory and models of social interaction, e.g. Bales' theory of Social Interaction Systems need to be translated into interaction issues in order to help to understand what is going on during a meeting and theories of verbal and nonverbal communication provide input for this understanding. Yet another important issue is the embedding of organizational information models that focus on the decision making process. The Issue Based Information Systems (IBIS) method is an attempt to support the decision making process by making the argumentation more explicit. Other high-level aspects that need to be taken into consideration are the organizational context of a meeting, the particular context (e.g., a project) of a meeting and the individual context for its participants.

## 2.5. Real-time, off-line and remote meeting assistance: Technology

In order to guide the research in this project several use cases have been developed. The general assumptions behind these use cases are that there is a need for being able to access available meeting information off-line. The information may be made available through a browsing facility and by asking questions about the meeting (a specific item of information, a summary, some global information, etc.). There is also a need to have real-time support during a meeting, i.e., being able to access information from previous meetings on the same topic or from a previous part of the current meeting. A third aim that needs to be mentioned is the possibility to take part in a meeting as a remote participant, where the meeting tools allow for the lack of available information (not knowing who plans to take the floor, not knowing who is currently speaking, not really being aware of something interesting going to

happen, etc.). The context for the use cases are the previously mentioned meetings of design teams. The use cases that have been introduced are about looking up information on previous meetings, auditing unattended meetings, reminding during meetings about contents of prior meetings and catching up on a meeting you are late for. Similarly, use cases have been developed for remote meeting assistance and for live meeting assistance. A meeting assistant can for example alert a participant that he still needs to give an opinion or that the matter under discussion contains some elements that are important for him. A lot of useful information that can be collected during a meeting can be presented to the chairman in order to guide his decisions. We will return to this meeting assistance in the next sections.

In order to collect and interpret data from a meeting in progress, a lot of multimodal recognition and interaction technology, based on models of verbal and nonverbal communication, multi-party interaction and group behavior in a meeting context has to be developed. In fact, this technology development takes up the main part of the current AMI project. Since this paper is not about the development of that particular technology, but rather about how to use it in augmented reality meeting environments, we confine ourselves here to mentioning the main research topics and, when necessary for our purposes, discuss more about the theory and technology in later sections that zoom in on augmented reality meetings. Hence, a short list of topics is presented here:

- Multimodal source localization, tracking, participant and speaker identification;
- Recognition of speech, gestures, postures, facial expressions, and emotions;
- Fusion (integrating information obtained from different media sources) and fission (selecting information and multi-media for information presentation);
- Automatic identification and modeling of conversational roles (speaker, hearer, addressee, audience, etc.); recognition of individual behavior using verbal and nonverbal cues;
- Detecting and modeling dialogue acts, turn-taking behavior, and focus of attention; detecting of argumentative structures in meetings, detecting of topics and topic shifts, detecting of decision points;
- Segmentation of multimodal streams, structuring by meeting events, identification of group activity;
- Technology for access to meeting information (retrieval, filtering, browsing, multimodal summarization, visualization, replay);
- Design of (real-time) meeting support, design of intelligent and pro-active meeting assistants, allowing remote and virtual presence, mixed reality and virtual reality tools and meeting environments.

A few additional remarks are in order. First of all, to make this research possible, data has to be collected and

environments need to be created (smart meeting rooms) that allow the collection of meeting data. This has been mentioned in section 2.2. For the data that is collected we design and develop theory, algorithms and tools that allow for automatic, semi-automatic and manual annotation. This has been discussed in section 2.3. In order to do so, it is useful to have knowledge about meeting processes and group interaction and behavior (section 2.4).

Again, as mentioned above, research in these areas is guided by a collection of use cases. Obviously, when translating research results into meeting support technology we can distinguish different levels of functionality in the technology (e.g. the intelligence of meeting assistants) and therefore different levels of intelligence to be obtained from the research that aims at automatic annotation and interpretation. Current state-of-the-art research covers the areas that are mentioned and also partial results from different areas can be integrated into representations and interpretations on which the development of useful tools for on-line and off-line meeting support can be based.

### 3. On- and off-line support for interpreting activities

Theory, models and algorithms that describe multi-party interaction (in our research, in particular during meetings) allow the design of tools and environments that support such interaction. We distinguish:

- Interpretation of events in the observed environment;
- Providing real-time support to activities in the observed environment;
- On and off-line multimedia retrieval, reporting (filtering), browsing and other ways of presentation;
- On-line observation and participation in activities;
- Owning and controlling the environment (and its inhabitants)

When looking at these issues, there is no need to confine ourselves to (smart) meeting rooms. Points of view and technology to be obtained can be applied to smart office environments, to educational environments, to home environments, and to public spaces. Depending on the point of view and the environment, more or less attention can be paid to issues of efficiency, privacy, control, ownership of access and information, trust, presence, well-feeling, family-feeling, social relationships, entertainment, and education. One other issue that should be mentioned is the role of autonomous and semi-autonomous (embodied) agents. This role will be discussed further in the forthcoming sections.

It is not the aim of this paper to discuss all the issues mentioned in the bullets above. We will focus on providing real-time support to activities in the observed environment and we will look at issues related to on-line observation and participation in activities.

Participants in activities, whether they are in a joint physical space or participate remotely, can be supported by the environment in their activities and this support can be realized by introducing virtual agents in the environment that act as assistants. For example, meeting assistants may have knowledge about previous meetings or about a current meeting. They can be available for all participants, they can act as a personal assistant to one particular meeting participant, or they can take part or all of the responsibility for the organization and the success of the meeting. Obviously, the latter needs to be done in cooperation with a human chairperson responsible for the organization and the outcome of the meeting.

In order to provide this support, we can introduce software agents acting in the meeting environment, collecting and interpreting information that has to be derived from verbal and nonverbal interactions between meeting participants (including interactions with smart objects, the environment, virtual and embodied agents, etc.). For example, a global agent that is allowed to communicate with the chairman as its personal assistant, collecting statistics about the talkativity, the dominance or other signs of involvement of the meeting participants. Also interesting is a situation where the agent assisting the chairman has some kind of knowledge about the emotional states of the participants, their preferences and their reasoning taking place during a decision-making process, in order to guide the meeting to a jointly accepted result.

Meeting assistants have been the topic of research in various projects (see e.g., the Neem project [3]). They can take the form of content assistants that know about finding and presenting related information from previous meetings and other information sources, organizational assistants that take responsibility for planning the meeting (negotiating time and place), meeting preparation (room reservation, prepare data projector and set-up of presentation, advise on time constraints during meeting), and remote control assistants (taking care of automatic slide changes during a presentation, dimming the lighting when presentation starts, etc.).

An other distinction, i.e. a distinction between agents that are embodied and agents that show themselves using other representations (prompts, menu's, question and answer forms, etc.) will be discussed in the next section. We will return to meeting assistants in section 4.4 where we discuss them in the context of our research.

#### 4. Meetings in the virtuality continuum

Localization and tracking technology and technologies for recognition and interpretation of verbal and nonverbal meeting activities allow real-time support of human activities taking place in a meeting environment. Since the technologies do not cover the recognition and interpretation of all aspects of activities we should also understand

that support that can be derived from this recognition and interpretation can only be available on a similarly limited level of functionality. This level will increase when theory, models and technology improve. Real-time support allows transformations from the environment's multimodal input to representations that can be mapped on multimedia output that is useful for the meeting participants, whether they are present in the smart environment or whether they are participating remotely. Obviously, during this transformation the information that is obtained from the input that is received can be enriched, where the enrichment employs knowledge available from the history of activities and interactions, knowledge obtained from ontologies underlying the domains of discussion, common sense knowledge, and, finally, knowledge about the participants and their current preferences.

One possible representation that can be obtained this way is a virtual reality representation of the meeting and the meeting activities. That is, there is a (preferably real-time) transformation from events and interactions in the physical meeting room to a virtual representation of events and interactions in a virtual meeting room. Being able to do so and the usefulness of doing so is discussed below. We mention that this idea fits into our earlier observations on sharing physical and virtual spaces and they fit also in general observations on augmented and mixed reality [7].

Since we need to be able to represent humans in the virtual meeting environments we need to discuss their virtual replica's (avatars, virtual humans, embodied conversational agents). They can be replica's of actual persons, but it is of course also possible to design embodied agents that have specific tasks in the environment and that communicate with actual persons or replica's of actual persons. Before discussing possible roles of virtual humans in virtual meeting environments we will consider two questions.

- Why do we want to have a virtual representation?
- How can we obtain a virtual representation?

Once we have discussed these issues we will turn to roles that can be played by the virtual humans in the virtual meeting environment.

##### 4.1. Why do we want to have a virtual representation?

There are several reasons to be interested in realizing a transformation from meeting activities to their virtual reality representations and in realizing a virtual meeting room (see also [10]).

- First of all, this transformation allows a 3D presentation and replay of multimedia information obtained from the capturing of a meeting. Depending on the state of the art of speech and image processing (recognition and interpretation) one may think of manual

annotation replay, replay based on both manually and automatically obtained annotations and interpretations and replay purely based on fully automatically obtained interpretations. Obviously, when the meeting environment has the intelligence to interpret the events in the meeting environment, it can transform events and present them in other useful ways (summaries, answers to queries, replays offering extra information, visualization of meta-information, etc.);

- Secondly, transforming annotations, whether they are obtained manually or automatically, can be used for the evaluation of annotations and annotation schemes and of the results obtained by, for example, machine learning methods. Current models of verbal and non-verbal interaction, multi-party interaction, social interaction, group interaction and, in particular considering our domain of meeting activities, models of meeting behavior on an individual or on a group level, are not available or only available for describing rather superficial phenomena of group interaction. Our virtual room offers a test-bed for eliciting and validation of models of social interaction, since in this representation we are able to control the display of various independent factors in the interaction between meeting participants (voice, gaze, distance, gestures, facial expressions) and therefore it can be used to study how they influence features of social interaction and social behavior.
- Thirdly, a virtual reality environment can be used to allow real-time and natural remote meeting participation. In order to do so we need to know which elements of multi-party interaction during a meeting need to be presented in a virtual meeting in order to obtain as much naturalness as possible. The test-bed function of a virtual meeting room, as mentioned above, can help to find out which (nonverbal) signals need to be mediated in one or other way.

#### 4.2. How can we obtain a virtual representation?

Information needed to build a virtual representation of meeting activities can be obtained in real-time from recordings of behaviors in real meetings (e.g. tracking of head or body movements, voice), from manual annotations or from machine learning models that induce higher level features from recordings and annotations. Obviously, when the main part of the annotations are obtained manually by annotators that off-line annotate the meeting, generation or presentation of this meeting information can become close to being perfect. This assumes that the annotation schemes that are used by human annotators are sufficiently detailed to allow (re-)generation of verbal and nonverbal behavior of virtual meeting participants.

The more complete the automatic annotation can be, the more complete a real-time regeneration in virtual re-

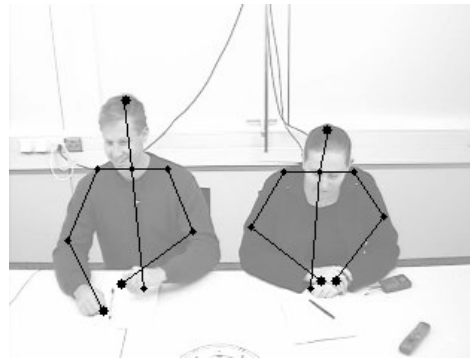


Figure 2. Pose extraction

ality can be. We used image extraction techniques (silhouette extraction, tracked and labeled skin regions) for human pose and gesture recognition. That is, algorithms are used to automatically extract poses and gestures of meeting participants and presenters from video recordings. The aim is to replay their behavior in animations on 3D avatars in our virtual meeting room environment. Therefore we map the perceived poses, gestures and estimated joint rotations on 3D avatar skeleton poses and gestures (Figure 2). This can be done in real-time resulting in reasonable convincing animations presented on H-Anim avatars (Figure 3). In addition we have looked at head orientation as some kind of substitute for gaze behavior. In Figure 4 we display a meeting where the participants are equipped with flock-of-birds sensors to capture their head movements during a meeting. In Figure 5 the transformation, using automatic pose and gesture recognition and recognition of head orientation, to the virtual meeting room is shown.

This real-time recognition allows teleconferencing, that is, real-time participation by remote participants. Similarly important, it also allows to simulate what humans did in the past, represent what humans do in a remote meeting and it allows adding virtual and embodied meeting assistants. It is not the case that we have designed a fully immersive virtual space for remote participants. Neither have we considered yet converting the in-



Figure 3. Pose and gesture animation

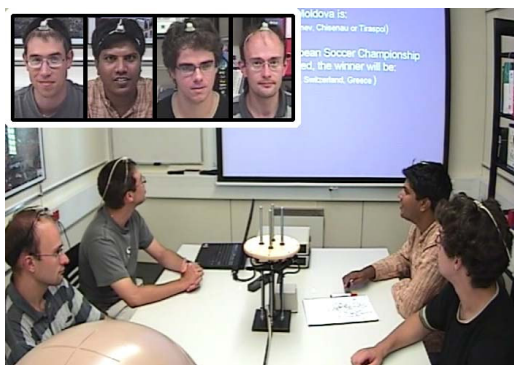


Figure 4. Flock-of-birds for head orientation

formation that is captured from individual cameras in front of each participant. Hence, presently there are no individual characteristics in the appearances, facial expressions don't show and emotion display and gaze behavior is not available.

Intelligence, based on recognition and interpretation of meeting events, is necessary for making an environment smart, e.g. by introducing intelligent meeting assistants, and by adding intelligence to agents that inhabit the virtual room (see next section). This is where the annotations of section 2.2 enter the play and why we need to improve the algorithms that aim at real-time recognizing (first-level annotations) and interpreting (second-level annotations) of verbal and nonverbal interactions during meetings. A virtual room obtained this way can further be augmented with statistics and visualizations and tuned to user preferences.

### 4.3. What roles do virtual humans play?

Virtual meeting participants can mimic what is happening in the physical meeting room. This is done with our transformations. There is not always a need to mimic everything and moreover technology or real-time constraints do not necessarily allow that. On the other hand, it is also possible to add information to the behavior of a virtual participant in order to improve communication. Certain characteristics in the behavior, e.g., gaze behavior to smoothen turn-taking or making clear who is addressed, can be added. The latter requires automatic addressee detection if we want to do this in real-time. Obviously, a remote meeting participant can choose to send his avatar to a meeting displaying only listening behavior [5]. The participant is continuously represented, but only when necessary its

owner is alerted and takes part, e.g. in a mimic mode.

### 4.4. Meeting assistants revisited

In section 3 we discussed the introduction of meeting assistants, agents that know about certain aspects of a meeting and that assist the various participants (including e.g. a chairman, a note-taker, an off-line visitor and a remote participant). Presently our work focuses on two of these meeting assistants. One is a virtual chairman, an assistant that during a meeting gathers information that is useful for a chairman; in fact, it may act as a chairman. Guarding agenda and time constraints is an obvious task. This also means taking care of the decision-making process; try to exploit the expertise of the meeting participants, decide about a presentation, etc. Active software agents assisting a meeting are discussed in [3]. Dominance detection is one of the topics that we research in order to assist a virtual chairman [12]. The second meeting assistant we focus on is an embodied presenter [8]. In the remainder of this section we discuss our work in progress on this presenter.

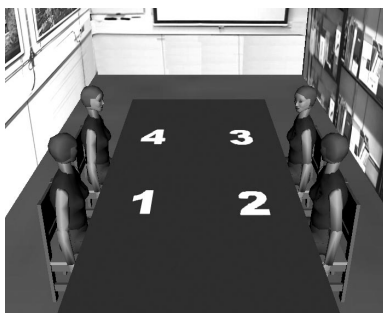
When discussing virtual presenters we can look at on-line presentations where the virtual presenter mimics a human presenter (and maybe add some characteristics in order to improve the performance), presentations where meeting participants present their (power point) sheets



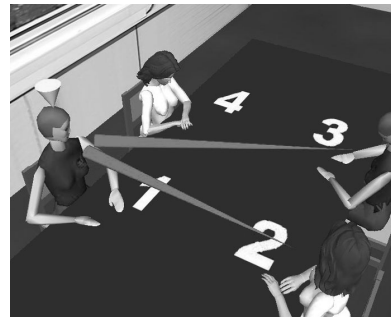
(a) Real setting of recorded meeting



(b) VMR seen by a participant



(c) VMR central view



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

Figure 5. Real and virtual meeting room



Figure 6. Virtual presenter in the VMR

with the help of their presentation agent or off-line presentations where an earlier presentation is regenerated. At this moment, our main concern is to model an embodied agent, and we delay questions that are related to off-line, on-line and (semi-) autonomous behavior to the future. We look at existing presentations, available from the AMI project, and try to design a script language to represent a presentation as a number of synchronized expressions, that is, create an example presentation script and rather natural presenter characteristics that allow us to replay an existing presentation in virtual reality. Sheet control, pointing gestures and speech are among the first modalities of a presentation that need to be modeled.

Pointing and gestures are other main issues of research for the virtual presenter. Pointing occupies a separate channel in the synchronization language. Apart from constraints on gesture phases (preparation, stroke and retraction) there are synchrony rules that need to be implemented and that take into account the phonological, syntactic, semantic and other rules (e.g. turn taking) that guide the behavior of a virtual presenter. Other issues that are addressed are gaze, hand shape (when pointing) and timing of pointing (moving from a start point to a target area on a sheet). In the future we will look at other gestures and posture shifts that can be employed by a presenter, and, most importantly, the possibility to interrupt a presenter [8]. See also Figure 6.

## 5. Conclusions and future work

We discussed how to go from captured data in a smart meeting room situation to a virtual meeting room. We discussed what needs to be annotated, where in an ideal situation the annotations are obtained automatically and in real-time converted to virtual reality generation, but in practice there is a mix of manually and automatically obtained annotations not allowing real-time generation) or there is real-time generation based on imperfect and incomplete data. Whether the latter is a problem depends on

the application. Our research is part of the European AMI project. In this project much more is captured (e.g. emotions) than we have taken into account when moving from a physical meeting room to a virtual meeting room. In future extensions of this project we hope to be able to extend our research to such issues and to the introduction of more intelligent and interactive virtual meeting participants (assistants) than we have considered here.

**Acknowledgements:** We described joint work of the Human Media Interaction (HMI) group of the University of Twente. This work was partly supported by the EU 6th FWP IST Integrated Project AMI (Augmented Multi-party Interaction, FP6-506811, publication AMI-129).

## References

- [1] R.F. Bales. Social Interaction Systems: Theory and Measurement. Transaction Publishers, NB, USA, 2001.
- [2] J. Carletta, S. Evert, U. Heid, J. Kilgour, J. Robertson & H. Voormann. The NITE XML toolkit: flexible annotation for multi-modal language data. Behaviour Research Methods, Instruments, and Computers 35 (3), 2003, 353-363.
- [3] C. Ellis & P. Barthelmess. The Neem dream. TAPIA '03, October 2003, Atlanta, Georgia, USA, 23-29.
- [4] C.M. Greenhalgh and S.D. Benford. Virtual reality teleconferencing: Implementation and experience. In Proc. 4<sup>th</sup> ECSCW, September 1995.
- [5] R. M. Maatman, J. Gratch and S. Marsella. Responsive Behavior of a Listening Agent. ICT TR 02.2005, 2005.
- [6] J.E. McGrath. Groups: Interaction and Performance. Prentice Hall, Englewood Cliffs, N.J., 1984.
- [7] A. Nijholt. Gulliver Project: Performers and Visitors. In: Proc. EVA 2002: "Electronic Imaging & the Visual Arts". V. Cappellini et al. (eds.), Pitagora Editrice Bologna, 241-245.
- [8] A. Nijholt, H. van Welbergen & J. Zwiers. Introducing an Embodied Virtual Presenter Agent in a Virtual Meeting Room. In: Proc. IASTED Intern. Conf. on AI and Applications (AIA 2005), M.H. Hamza (ed.), Acta Press, Anaheim, 2005, 579-584.
- [9] A. Nijholt, R. op den Akker & D. Heylen. Meetings and Meeting Modeling in Smart Environments. AI & Society, J. of Human-Centred Systems, Springer London Ltd, 2005, SpringerLink On-line version, August 2005.
- [10] D. Reidsma, R. op den Akker, R. Rienks, R. Poppe, A. Nijholt, D. Heylen & J. Zwiers. Virtual Meeting Rooms: From Observation to Simulation. CD Proc. Social Intelligence Design 2005, Stanford University, USA, R. Fruchter (Ed.).
- [11] D. Reidsma, N. Jovanovic & D. Hofs. Designing annotation tools based on properties of annotation problems. TR-CTIT-04-45, University of Twente, November 2004.
- [12] R. Rienks & D. Heylen. Automatic dominance detection in meetings using support vector machines. ICMI-MMMP Workshop, Trento, October 2005.