

Mixed reality participants in smart meeting rooms and smart home environments

Anton Nijholt · Job Zwiers · Jan Peciva

Received: 25 July 2006 / Accepted: 15 September 2006 / Published online: 25 April 2007
© Springer-Verlag London Limited 2007

Abstract Human–computer interaction requires modeling of the user. A user profile typically contains preferences, interests, characteristics, and interaction behavior. However, in its multimodal interaction with a smart environment the user displays characteristics that show how the user, not necessarily consciously, verbally and nonverbally provides the smart environment with useful input and feedback. Especially in ambient intelligence environments we encounter situations where the environment supports interaction between the environment, smart objects (e.g., mobile robots, smart furniture) and human participants in the environment. Therefore it is useful for the profile to contain a physical representation of the user obtained by multi-modal capturing techniques. We discuss the modeling and simulation of interacting participants in a virtual meeting room, we discuss how remote meeting participants can take part in meeting activities and they have some observations on translating research results to smart home environments.

Keywords Smart environments · Ambient intelligence · Embodied agents · Remote participation · Virtual reality

1 Introduction

Human–computer interaction requires modeling of the user in the interface. User modeling has become a well-respected research area and knowledge about the user makes it possible for a system to adapt its behavior towards the user, e.g. by predicting the user’s behavior and preferences and anticipating on this behavior and preferences. There is a tendency to collect as much information of a user as possible. A user profile typically contains preferences, interests, characteristics, and interaction behavior. During the interaction with a system a user displays behavior and makes decisions that can be used to modify a profile. During the interaction it is however more important that the system knows about details of the needs of the user at that particular moment than the global information that is available in a user profile.

During multimodal interaction a system has the possibility, using multiple sensors, to capture real-time the changing characteristics of the user and its way of interacting. This may include facial expressions, gestures, intonation, body posture and biometric information. Fusion and interpretation of that information will make it possible to decide whether a user is satisfied or frustrated about what is going on in the interaction. We have a real-time modeling of the user. It is certainly not the case that for all human–computer interaction this real-time modeling of the user is required and useful. On the other hand, there are applications for which we need to go several steps further. In smart environments or ambient intelligence environments we encounter situations where the computerized environment has to support interaction between the environment, smart objects (e.g., mobile robots, smart furniture) and human visitors or inhabitants of the environment.

A. Nijholt (✉) · J. Zwiers
University of Twente, CTIT, PO Box 217,
7500 AE Enschede, The Netherlands
e-mail: anijholt@cs.utwente.nl

J. Zwiers
e-mail: zwiers@cs.utwente.nl

J. Peciva
Faculty of Information Technology,
Brno University of Technology,
61266 Brno, Czech Republic
e-mail: peciva@fit.vutbr.cz

This situation is not really different from a situation where users become part of an augmented reality or virtual reality environment and the environment needs to know about or be able to capture movements and body properties of a user of that environment. Since we are talking about multiple interacting human users or visitors of these interaction supporting environments the question is how to represent these users of such environments. The user profile may contain a physical representation of the user and multi-modal capturing techniques may add in real-time dynamic changes (movements, facial expressions, posture shifts, gestures, etc.). Obviously, the need to present this information to other users in the environment is higher in a situation where users share a virtual environment and one or more of them are not physically present, than in a situation where they share the same physical environment.

In this paper, we discuss the modeling and simulation of interacting participants in a smart meeting environment and we have observations on how to translate research results obtained in the meeting domain to other domains, in particular the domain of smart home environments.

The organization of this paper is as follows. In Sect. 2 we have some general observations on extensions of, more or less, traditional ways of user modeling. That is, we look at users—or rather visitors, partners, collaborators, colleagues, inhabitants, etc.—acting in smart and virtual environments for which it is useful to include in a profile properties dealing with location preferences and behavior, properties dealing with physical (appearance) and other observable characteristics of verbal and nonverbal behavior. In Sect. 3 we zoom in on teleconferencing and how work in this area is related to several European and DARPA funded projects on meeting modeling. Sects. 4 and 5 show our application of virtual and distributed virtual meeting rooms where meeting participants are represented by virtual humans. Section 6 of this paper contains observations on why this research is relevant for real-time support in smart home environments and in Sect. 7 we present conclusions.

2 Modeling partners, participants and inhabitants

User profiles allow computer users to be presented with personalized applications. Typically a profile contains preferences, interests, characteristics and behavior. Much more can be added, but in traditional human–computer interaction there is not always a need to process that information. When the system the user interacts with allows multimodality then more information about the user can be extracted in real-time. For example, the system may learn about interaction pattern preferences [1] or detect the user's emotional state and adapt its interaction behavior, its

interface and its feedback according to them. The body and what the user is doing with his or her body is becoming important for the system and this is even more the case when the user is allowed to move around and interact from different positions and with various objects, maybe other users and parts of a computer-supported or monitored environment. We not only have users, but also inhabitants, players, partners and passers-by. Not only they need to be characterized, but they need to be characterized in their physical context from information obtained from sensors in the environment and its objects (location sensors, cameras, tracking systems, microphones) including wearables, portable devices and active and passive tags attached to the users. Rather than interaction histories these perceptual technologies allow us to build up and exploit context histories [2].

In ambient intelligence research the aim is to model verbal and nonverbal communication and other human behavior in such a way that the environment in which this communication and other behavior takes place is able to support these human activities in a natural way.

Obviously, the purposes of the environments and the aims of the inhabitants of a particular environment can very much constrain and guide the interpretation of the activities and the support given by the environment.

Entertainment, education, profession, home, family, friends, etc., all provide different viewpoints on activities, communications, and desirable real-time support and sometimes also on off-line support allowing intelligent access to archived activities and multi-media presentation of such information.

3 Supporting meetings and meeting partners

We start this section with five observations on teleconferencing.

- (1) There is a growing need for teleconferencing;
- (2) current, commercially available teleconferencing systems are hardly used;
- (3) current teleconferencing systems are very much biased towards transmitting video and do not consider other ways of transmitting participants' contributions, including manipulating their contributions, let alone, providing means to offer meta-information about the conference or meeting;
- (4) current teleconferencing systems assume that all conference participants are remote, rather than assuming that there can be several people in the same location taking part in the conference, and, finally,
- (5) current teleconferencing systems do not make use or anticipate research results in the areas of image

processing, artificial intelligence, animation, virtual reality and information visualization.

Obviously, these observations also hold when we look at web casting, remote viewing and audio/web conferencing. Here the emphasis is on offering the viewer advanced viewing facilities (panoramic views, speaker image, whiteboard and sheet views), although also here we see attempts to introduce interactivity. Also automatic camera and microphone control based on speaker localization or viewers' interests (e.g., made explicit by his gaze [3] is considered. However, in general there is poor media richness, interrupted media, delay in media delivery, and, in particular, lack of interactivity. Lack of interactivity means lack of engagement and a poor sense of presence [4].

The situation is slightly different when looking at Computer Supported Collaborative Work (CSCW) systems. The comparison is not completely fair, because here, from the beginning research issues were much more advanced since workers are assumed to collaborate in non-verbal ways (sharing notes, sharing objects), and therefore it is an advantage to have their actions made visible for each other and have a virtual environment designed that supports these activities, while the traditional viewpoint of meetings is that only the verbal interaction is important and needs to be captured. Joint virtual workspaces allowing access from 'remote' places and offering tools for designers and scientists to design and experiment are the future workspaces.

Hence, it is obvious to expect research on smart environments and ambient intelligence earlier to be associated with computer supported collaborative work than with teleconferencing. Research on smart environments and ambient intelligence is about capturing information various kinds of sensors (audio, video, motion and location sensors, wearables, etc.) about activities in an environment, interpreting and enriching this information, and making it available to inhabitants or virtual agents informing and guiding inhabitants. When we speak of inhabitants, we include situations where the environment is virtual and there is only computer-mediated contact between the inhabitants and situations where several people are in the same physical environment and others are allowed to enter this environment, and be virtually present, from remote places. Here, with virtual we do not necessarily mean virtual reality.

EU and DARPA funded research projects on multi-modal interaction have been designed to provide the link between smart environments and ambient intelligence research and meeting or teleconferencing research. EU projects are AMI (Augmented Multi-party Interaction) [5], CHIL (Computers in the Human Interaction Loop) [6] and AMIGO (Ambient Intelligence for the Networked Home

Environment) [7, 8]. The main DARPA funded project on small group meetings is CALO (Cognitive Agent that Learns and Organizes) [9].

The research reported in this paper grew out of the AMI project. The AMI project is a comprehensive research effort on modeling multi-party interactions in the context of meetings. Multi-party interactions are multimodal of nature; hence, multimodal interactions between meeting participants are subject of research. Group dynamics, group interaction, goals and aims of group members, current verbal and nonverbal meeting interaction, emotions, speech, gestures, poses and facial expressions need to be modeled in order to allow recognition and interpretation. This recognition and interpretation is needed to allow off-line access, but also to allow real-time support of activities by the meeting participants.

While in the AMI project the main starting point was the off-line browsing of meeting information, in the research reported here the emphasis is on

- (1) using AMI technology for real-time visualizing of interpreted meeting information, and
- (2) using network technology to give real-time access to this information.

In our distributed virtual meeting room (DVMR) experiments we have confined ourselves to a real-time representation of meeting events in an environment inhabited by embodied agents representing meeting participants. However, this is only one way to have a real-time mapping and transformation from meeting events to a remotely accessible multimedia representation that allows remote participation, remote experiencing and remote access to meta-information of a meeting. In an off-line situation more effort can be given to the interpretation of the meeting data and to efforts to allow access to the data in such a way that the meeting in the past can nevertheless be experienced by an off-line 'meeting participant'.

In the next sections we have an overview of technology that has been developed to perform our distributed virtual meeting room experiments. It shows one particular way of connecting smart meeting environments, where each environment can have a number of inhabitants (meeting participants) or just one inhabitant (meeting participant).

4 Designing a virtual meeting room

4.1 From meeting events to multimedia representations

To get closer to our objectives we have looked at mapping meeting events to representations of these events, possibly enriched with meta-information about the meeting, in 3D virtual reality environments. In previous papers [10, 11]

(see also Sect. 4.3) we discussed how to obtain enriched virtual reality representations of meeting events from annotated meeting data, where part of the annotations could be obtained automatically and in real-time and where part of the annotation needed to be done manually. Obviously, the real-time obtained meta-information can be made accessible real-time, while the more comprehensive knowledge, obtained by integrating automatically and manually obtained information, can only be made accessible off-line.

4.2 Visualizing meetings and meeting events

Comprehensive interpretation of meeting interaction is far from being possible, it would require comprehensive interpretation and computational modeling of human individual and group behavior. Nevertheless, there is a level of available speech and image processing techniques that allows us to map captured (through microphones and cameras) meeting events (verbal and nonverbal interaction, identifying participants, and tracking of participants in the meeting environment) to multimedia online and off-line presentations of these events.

We have looked at transforming meeting events to events in a virtual reality representation of a meeting environment where embodied agents play the role of meeting participants. Given the limitations of real-time speech and image processing techniques, our main interest has been the mapping of the nonverbal behavior of human meeting participants to the nonverbal behavior of their representations as embodied agents in a virtual meeting environment. Being able to do this is a prerequisite of further and more intelligent processing of meeting information, including real-time access to meeting data and real-time participation in meetings.

4.3 Capturing meeting activity

In our research we have looked at capturing meeting activities from an image processing point of view and at capturing meeting activities from a higher-level point of view, that is, a point of view that allows, among others, observations about dominance, focus of attention, addressee identification, and emotion display. We will return to these issues in forthcoming sections, but here we will look at capturing a limited selection of nonverbal meeting interactions (posture, gestures, and head orientation) only, and we look at possibilities to transform them to a virtual reality representation of a meeting room and its meeting participants.

In order to capture nonverbal activities of meeting participants we studied posture and gesture activity, using our vision software package. Our flock-of-birds software

package was used to track head orientation of some of our four party meetings.

The computer vision software processes low resolution, monocular image sequences from a single camera. A silhouette is extracted, shadows are removed and skin color is extracted from the silhouette in order to locate hand and head. Silhouette matching is used to match a projection of a human body model to the extracted silhouette. This allows us to display animated representations of meeting participants in a (3D) virtual reality environment. The 3D positions of head, elbows and hands can reasonably be calculated [12]. 3D technology based upon portable standards, like VRML/X3D and H-Anim avatars is used. For some meetings to be recorded electromagnetic sensors were mounted on the heads of the participants for tracking their head movements. Especially in meetings this allows us to record and real-time display head orientations of the represented meeting participants. Although there can be differences in head orientation and gaze direction, it nevertheless allows a sufficiently realistic representation of focus of attention behavior (addressing persons, looking at a speaker, looking at notes or looking at the white board in the meeting room).

4.4 A virtual meeting room representation

The research described in the previous subsections allows us to design a virtual meeting room (VMR) in which the activities of human meeting participants are represented. Virtual reality allows us to view the room from all possible angles, for example the viewpoint of a participant, and by means of a head mounted display we can become immersed.

Due to our limited number of capturing devices, but also because of imperfect capturing technology and corresponding algorithms, the representation of meeting events is far from perfect. A more perfect representation can be obtained if we are able to use other real-time and automatically obtained annotations of a meeting or, forgetting about real-time constraints, use manually, off-line obtained annotations. Annotations might include results of speech recognition, dialogue structure recognition, talkativity, movements, speaker localization, turn-taking, slide changes, etc. and they can real-time trigger changes (viewpoint changes, adding of metadata, etc.) in the VMR. As such it can play a useful role during a meeting, either for remote viewers or for the meeting participants themselves. We will return to that below. First, we distinguish the following useful applications of our VMR environment [13]:

- First of all, it 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 nonverbal 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 [14]. 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.

As mentioned in the first bullet, the VMR allows us to reconstruct a meeting, but when useful we can do it in a different way. Gestures can be exaggerated, pointing can be done such that it is better recognizable, speech can be improved, and we can even have different combinations of modalities than used in the real meeting. A view of the current VMR is displayed in Fig. 1.

5 Designing a distributed virtual meeting room

As mentioned, a VMR can be used to allow real-time and natural remote participation. Participation requires real-

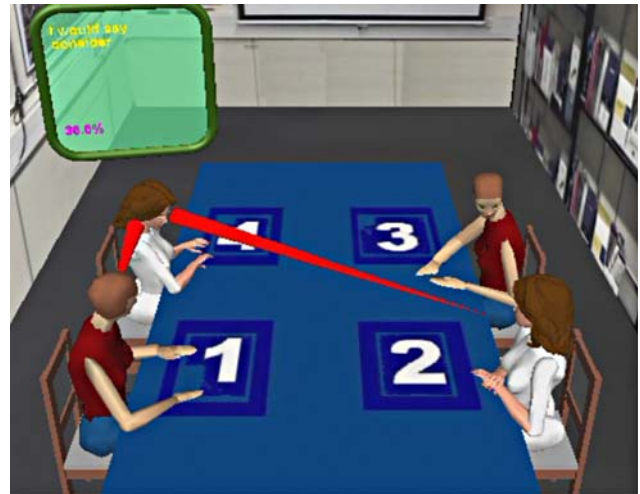


Fig. 1 The virtual meeting room showing gestures, head movements, the speech transcript, the addressee(s) of the speaker and the percentage of a person has spoken until that moment

time interaction with other meeting participants. This section is concerned with real-time use of the VMR.

5.1 VMR for life meeting assistance

Let us first consider the situation where we offer the VMR to the meeting participants inhabiting the physical meeting room while they are interacting. While meeting they can get all kinds of information about the meeting presented in this virtual environment and they can use it as a domain-dependent browser asking questions like: Who is this person, what did he/she say about this topic in a previous meeting, why is this person getting upset when we talk about this topic, etc. Hence, due to this visualization, meeting participants may feel stimulated to ask questions related to behavior of meeting participants, meta-information displayed in the environment and events taking place (without disturbing the meeting). Clearly, when looking at the VMR from this point of view it serves the role of providing life meeting assistance to the meeting participants present in the real meeting room. The visualization provides the context for the user to interact with the system and it provides the context for the system to interpret and assist the user.

Remote on-line viewing of the VMR is of course no problem. That is, non-meeting participants can get access and see what is going on. This does not require interactivity, although, inherently to virtual reality, any viewpoint can be taken, meaning that, e.g. the viewpoint from an empty chair at the table can be taken. This audience is not necessarily visible by the meeting participants. A slight extension should allow visualization of the audience, for example as avatars, and making the meeting participants,

still assuming that they use the VMR as life meeting assistant, aware of who is in the audience. We have not done this yet, but it fits in a tradition of multi-user virtual environments, where in this case the multi-user environment can be constrained to a public gallery, not disturbing the meeting. Obviously, many other ideas common in multi-user virtual reality environments and distributed virtual environments, including the various ways of distribution of data and processes can be introduced here [15].

5.2 VMR for distributed meeting assistance

The general objective of our distributed virtual meeting room (DVMR) is that we have different smart locations, each equipped with cameras, microphones and probably other sensors. These smart locations are inhabited by one or more meeting participants that take part in a virtual meeting room in which all locations and their inhabitants join. That is, we connect smart meeting rooms, we can connect individual remote participants to smart meeting rooms and we can connect many individual remote participants to one joint virtual meeting room. From every location we need to capture meeting behavior of the inhabitants and make it available to the joint virtual reality meeting room that can be accessed by every meeting participant from every location. In Fig. 2 we have illustrated the situation where two smart meeting rooms are connected and the captured information is displayed in a virtual reality meeting room.

In our setup, demonstrated at the MLMI 2005 conference in Edinburgh, local constraints and resource limitations did not allow us to demonstrate the full potential of our technology. We confined ourselves to a situation where one remote meeting participant joined a meeting of three

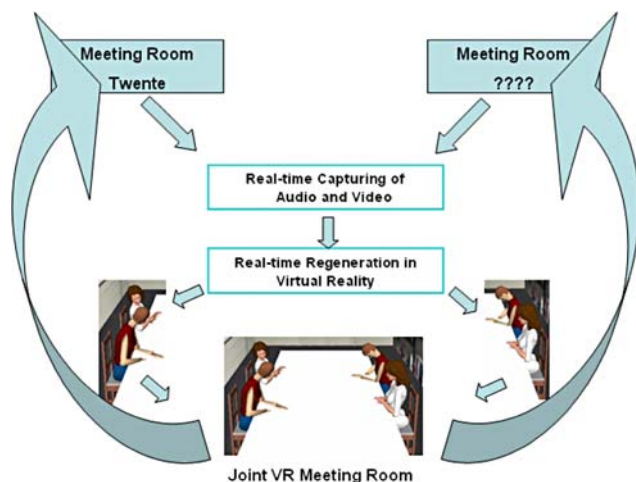


Fig. 2 Capturing and re-generation of meeting activities from remote locations

embodied agents in the form of an animated embodied agent in a virtual representation of the IDIAP smart meeting room. Capturing of the remote participant was done using a simple web camera and electromagnetic sensors to measure head orientation. In the near future we may expect that these latter sensors can be replaced by other, less non-obtrusive, sensors (e.g. rfid tags, glasses, headsets, wearables). At the moment we use technology developed in our group (vision software, flock-of-bird sensors, a multi-agent platform, and DVMR clients and server software) for tracking meeting activity in remote and connected environments and transforming it into activities displayed by embodied agents in a joint virtual reality meeting room. A remote participant, in fact every participant, can see the DVMR with avatars representing the meeting participants and can see the meeting activities of those participants.

The technology used within the DVMR experiment differs substantially from normal video conferencing technology. Rather than sending video data as such, this data is transformed in a format that enables analysis and transformation. For the DVMR experiment the focus was on representing poses and gestures, rather than, for example, facial expressions. Poses of the human body are easily represented in the form of skeleton poses, essentially in the same format as being used for applications in the field of virtual reality and computer games. Such skeleton poses are also more appropriate as input data for classification algorithms for gestures. Another advantage for remote meetings, especially when relying on small handheld devices, using wireless connections, is that communicating skeleton data requires substantially less bandwidth than video data. A more abstract representation of human body data is also vital for combining different input channels, possibly using different input modalities. Here we rely on two different input modalities: one for body posture estimation based upon a video camera, and a second input channel using a head tracker device. Although the image recognition data for body postures also makes some estimation of the head position, it turned out that using a separate head tracker was much more reliable in this case. The general conclusion is, not so much that everyone should use a head tracker device, but rather that the setup as a whole should be capable of fusing a wide variety of input modalities. This will allow one to adapt to a lot of different and often difficult situations.

In the long run, we expect to see two types of environment for remote meetings: specialized meeting rooms, fully equipped with whatever hardware is needed and available for meetings on the one hand side, and far more basic single user environments based upon equipment that happens to be available. The capability to exploit whatever equipment is available might be an important factor for the

acceptance of the technology. In this respect, we expect a lot from improved speech recognition and especially from natural language analysis. The current version of the virtual meeting room requires manual control, using classical input devices like keyboard or mouse, in order to look around, interact with objects etcetera. It seems unlikely that in a more realistic setting people that are participating in a real meeting would like to do that. Simpler interaction, based upon gaze detection but also on speech recognition should replace this situation.

The DVMR-Server transforms its input to an up-to-date distributed virtual meeting room. Objects in the DVMR can be controlled/moved by the DVMR inhabitants. As an example, since many of our recorded meetings are design meetings devoted to the design of a remote control, we designed a remote control and put it in the DVMR as an example of how real and remote meeting participants can discuss and manipulate the properties of this remote control. Clearly, visualizing and manipulating objects that are under discussion, whether they represent physical objects or documents and presentations, is an important issue in advanced meeting technology.

The remote participants have a virtual position at the table, and can watch the meeting from that virtual position or, if they prefer, can watch the meeting from a more global point of view. The same hold true for the other participants: they will see the remote person at his or her virtual position, making the movements and gestures of the real person. The technology is based upon simple consumer web cams, together with image recognition technology that extracts key features, like body position and gestures. This process is illustrated in Fig. 3.

Figure 4 shows that there is the possibility to transform meeting activities to other media, modalities and appearances before displaying them to meeting participants. We have chosen to make transformations from and to modalities, since that shows how detailed we can go; obviously, modality changes, changes of combinations of modalities and replacing human modalities by other media to present activities and information can be considered.

Each computer running the DVMR is transforming inputs from its input devices to its virtual meeting room replicas. Our distributed version of the VMR is using re-

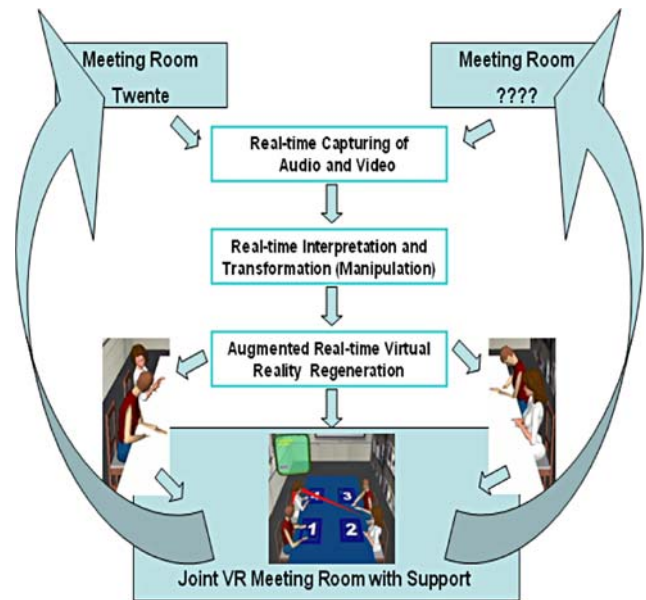


Fig. 4 Capturing, manipulation and re-generation of meeting activities from remote locations

cent developments in database technology based on delayed commits for time-stamped transactions in replicated multi-version databases. Objects in the DVMR can be controlled and moved by the DVMR inhabitants. As an example, since many of our recorded meetings are design meetings devoted to the design of a remote control, we designed a remote control and put it in the DVMR as an example of how real and remote meeting participants can discuss and manipulate the properties of this remote control.

The DVMR replicates meeting data among all participating computers. Three types of replication can be distinguished:

- static data
- primary-copy replication
- delta consistency replication

Static data are those that never changes. Once they are set at creation or loading time, they stay the same. No synchronization is necessary for these data.

Fig. 3 Remote participant, making some gesture, the gesture recognition, and the representation within the VMR, as seen through the eyes of one of the other participants



Primary-copy data includes, for example, avatars and all other objects that are modified just by one computer. Primary-backup replication [16] means that just the computer holding the primary replica can modify its value. All other computers hold “backups” that cannot be directly modified without contacting the primary replica.

The most complicated situation is when several participants want to modify one object simultaneously. For example, more people want to manipulate the remote control. Concurrent writes from different computers with different values may break the consistency of the scene, for instance, one participant might see a yellow remote control whereas others see a blue one. Therefore, some rules are necessary for definition and maintenance of consistency. Our solution is based on time-stamped transactions with delayed commits. It can be considered as extended delta consistency [17]. The concurrency rules grant that when concurrent writes happen, the same write will win over another write on all computers resulting in the same virtual scene databases on all computers, thereby ensuring consistency.

The concurrency rules are defined as follows:

- each write is causally dependent on previous write
- when more concurrent writes are found, the earliest of them is accepted

To get the rules working, global order of writes is necessary. Therefore each write is marked by globally unique timestamp. We are using Lamport Timestamps [18] without final acknowledging. Then, the global order of writes is established by sorting them based on their timestamps.

Because the order of writes cannot be established immediately, since it takes some time for updates to propagate through the network, we are using delayed commit. It means that the order of writes is considered temporary or speculative until writes are older than longest network latency between computers. Then, we are sure that no other older write will arrive to change the order of these old writes. The moment of turning write from “speculative” to “permanent” we call commit.

The concurrent writes are detected by the commit operation. If the write is about to commit and it does not depend on the last committed value, it has to be aborted since it depends on already overwritten value or other aborted write.

Dividing writes into two classes—speculative and permanent—makes it possible to do some optimizations related to network latency. For example, it is possible to change a color of the remote control speculatively and user can see the result immediately without waiting until network communication is done and the value is committed. This optimistic behavior has the advantage of showing an

immediate response to the user, while consistency of the scene is guaranteed by existence of “committed” scene. If any write that is still not committed turns out to break the scene consistency, it is aborted and its effect is removed from the “speculative” scene that is shown to the user.

5.3 Current research issues

Currently we work on the integration of speech recognition in the DVMR in order to select or manipulate objects or agents in the environment. For speech software we use SpeechPearl XML (ScanSoft). Another issue that is being looked at is capturing and mediating gaze behavior between remote meeting participants. Sharing and manipulating shared objects is another issue that is requiring our attention. Obviously, and also part of our current research efforts, is virtual reality visualization of meeting events, interpretations of these meeting events in order to produce semantic preserving transformations, and presentation of these meeting events using various media sources. We also hope to integrate current research efforts on personalization of embodied agents, facial expressions and emotion display into our efforts to obtain meeting environments that allow for more natural meeting experiences.

6 From smart meeting technology to smart home technology

In the domain of meetings supported by smart environment technology it is useful to provide support during the meeting, it is useful to allow people who can not be present to view what is going on, it is useful to allow people to remotely participate and it is useful to provide access to captured multimedia information about a previous meeting, both for people who were present and want to recall part of a meeting and for people who could not attend. These issues are also important and can be explained and explored in the context of smart home environments.

It should be clear that topics such as visualization, virtual reality, embodied agents (virtual humans) and remote participation can become important issues, assuming appropriately sensor equipped smart home environments, to support

- (1) Multi-party interaction and joint activities of family members (including mobile robots, virtual pets and virtual humans),
- (2) Real-time monitoring of activities and participation in such activities, and
- (3) Retrieving, browsing, and replaying of previously captured and stored information about activities that took place in a particular home environment.

Recording of family events, real-time sharing events with those that are not there, remote participation, presently often in primitive ways, and playing around with recorded material in order to re-experience previous events, all these activities take place nowadays and can be done in more intelligent, more creative and more entertaining way with smart home technology that resembles the technology we discussed in this paper [19]. Obviously, it should not be understood that people living in the same environment always will have the need to have their smart home environment turned on to perform all these tasks. For some tasks this will be the case (for example, control of energy consumption, preventing non-authorized access), for other tasks (for example, allowing virtual access to a personal mixed reality environment) more explicit decisions by the inhabitants will be needed. However, also in the latter case the issue of controlling, owning and maintaining the environment by others than the inhabitants remains.

7 Conclusions

From detecting rather straightforward events as entering a room, being in the proximity of a certain object or identifying a person in the room, to the interpretation of events in which more persons are involved is a rather big step. However, in ambient intelligence research small steps in this direction are taken. In this paper we focused on connecting different locations and visualizing the joint activity in one virtual room. In the context of meetings this allows connecting physically remote meeting rooms. It can also connect a single meeting participant travelling around or sitting in his or her own office environment to a smart meeting room located somewhere else. Clearly, other research has been done to realize these aims [20]. Ideas available in previous research have been extended in this paper, making use of research results that have become available in research projects on ambient intelligent and on multi-party interaction. In the context of smart home environments we can have travelling family members sitting in their hotel room connecting to home activities (join a dinner or a birthday party, virtually hugging a child when it is bedtime). We discussed some technical issues that allow us to regenerate a scene in the real world into a virtual reality representation. Based on some level of understanding scenes and events meta-information can be added to the virtual representation or the information can be manipulated in such a way that a more appropriate or enjoyable representation can be visualized. Based on a similar level of understanding scenes and events recorded information can be stored and made accessible for off-line retrieval or replay.

Acknowledgments This work was partly supported by the European Union 6th FWP IST Integrated Project AMI (Augmented Multi-party Interaction, FP6-506811, publication AMI-190).

References

- Oviatt S, Lunsford R, Coulston R (2005) Individual differences in multimodal integration patterns: what are they and why do they exist? In: CHI '05: Proceedings of the 2005 conference on human factors in computing systems, pp 241–249
- Prante et al. (2005) ECHISE, 1st International workshop on exploiting context histories in smart environments. Workshop at PERSASIVE 2005, Munich, Germany
- Smith D, et al. (2005) OverHear: augmenting attention in remote social gatherings through computer-mediated hearing. In: Proceedings CHI 2005, Portland, Oregon, USA, ACM, pp 1801–1804
- Lombard M, Ditton TB (1997) At the heart of it all: the concept of presence. *J Comput Mediated Comm* 3(2). Available at <http://www.ascusc.org/jcmc/vol3/issue2/lombard.html>
- McCowan I, Gatica-Perez D, Bengio S, Moore D, Bourlard H (2003) Towards computer understanding of human interactions. In: Aarts E et al. (ed) Ambient intelligence, Lecture Notes in Computer Science, Springer, Heidelberg, pp 235–251. <http://www.amiproject.org/>
- Waibel A, Steusloff H, Stiefelhagen R (2004) CHIL—Computers in the human interaction loop. In: 5th international workshop on image analysis for multimedia interactive services, April, Lisbon, Portugal. <http://www.chil.server.de/>
- CALO: <http://www.cse.ogi.edu/CHCC/Projects/CALO/>
- AMIGO: <http://www.amigo-project.org/>
- Rocker C, Janse MD, Portolan N, Streitz N (2005) User requirements for intelligent home environments: a scenario-driven approach and empirical cross-cultural study. In: Proceedings of the 2005 joint conference on smart objects and ambient intelligence: innovative context-aware services: usages and technologies, Grenoble, France, pp 11–116
- Reidsma D, Op den Akker R, Rienks R, Poppe R, Nijholt A, Heylen D, Zwiers J (2005) Virtual meeting rooms: from observation to simulation. In: Fruchter R (ed) Proceedings of social intelligence design 2005, Stanford University, CD Proceedings
- Rienks R, Nijholt A, Reidsma D (2006) Meetings and meeting support in ambient intelligence, Chap. 18. In: Vasilakos ThA, Pedrycz W (eds) Ambient intelligence, wireless networking, ubiquitous computing. Artech House, Norwood
- Poppe R, Heylen D, Nijholt A, Poel M (2005) Towards real-time body pose estimation for presenters in meeting environments. In: Skala V (ed) Proceedings of 13th international conference in central Europe on computer graphics, visualization and computer vision. Plzen, Czech Republic, pp 41–44
- Nijholt A (2005) Meetings in the virtuality continuum: Send Your Avatar. In: Kunii TL, Soon SH, Sourin A (eds) Proceedings 2005 international conference on CYBERWORLDS, IEEE Computer Society Press, Los Alamitos/Singapore, pp 75–82
- Traum D, Rickel J (2002) Embodied agents for multi-party dialogue in immersive virtual worlds. In: Proceedings of the first international joint conference on autonomous agents and multi-agent systems (AAMAS 2002), July, Bologna, Italy, pp 766–773
- Saar K (1999) VIRTUS: a collaborative multi-user platform. In: Proceedings of the fourth symposium on Virtual reality modeling language. Paderborn, Germany, pp 141–152
- Wiesmann M, Pedone F, Schiper A, Kemme B, Alonso G (2000) Understanding replication in databases and distributed systems. In: Proceedings of ICDCS'2000, Taipei, Taiwan, ROC, pp 264–274

17. Singla A, Ramachandran U, Hodgins J (1997) Temporal notions of synchronization and consistency in Beehive. In: Proceedings of the ninth annual ACM symposium on parallel algorithms and architectures (Newport, Rhode Island, United States, June 23–25), SPAA '97. ACM Press, New York, pp 211–220
18. Lamport L (1978) Time, clocks and the ordering of events in a distributed system. *Commun ACM* 21(7):558–565
19. Gamhewage C, de Silva, Byoungjun Oh, Toshihiko Y, Kiyoharu A (2005) Experience retrieval in a ubiquitous home. In: International multimedia conference. Proceedings of the 2nd ACM workshop on continuous archival and retrieval of personal experiences, Singapore, pp 35–44
20. Frecon E, Nou AA (1998) Building distributed virtual environments to support collaborative work. *VRST '98*, November, Taipei, Taiwan, pp 105–113