Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Information Sciences 178 (2008) 612–630
INFORMATION SCIENCES
AN INTERNATIONAL JOURNAL
www.elsevier.com/locate/ins

Google home: Experience, support and re-experience of social home activities ☆

Anton Nijholt

Human Media Interaction, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

Abstract

Ambient intelligence research is about ubiquitous computing and about social and intelligent properties of computer-supported environments. These properties aim at providing inhabitants or visitors of ambient intelligence environments with support in their activities. Activities include interactions between inhabitants and between inhabitants and (semi-) autonomous agents, including mobile robots, virtual humans and other smart objects in the environment. Providing real-time support requires understanding of behavior and activities. Clearly, being able to provide real-time support also allows us to provide off-line support, that is, intelligent off-line retrieval, summarizing, browsing and even replay, possibly in a transformed way, of stored information. Real-time remote access to these computer-supported environments also allows participation in activities and such participation as well can profit from the real-time capturing and interpretation of behavior and activities performed and supported by ambient intelligence technology. In this paper, we illustrate and support these observations by looking at results obtained in several European and US projects, in particular projects on smart environments, whether they are smart meetings or lecture rooms, smart offices or intelligently monitored events in public spaces. In particular, we look at the augmented multi-party interaction (AMI) project in which we are involved and we try to sketch a framework in which we can transform research results from the meeting context to the home environment context.

Keywords: Ambient intelligence; Smart home environments; Multi-modal interaction; Multi-party interaction; Experience capturing; Multi-media retrieval and browsing

1. Introduction

There is Google and there are other search engines. They are designed – and also their further development can be foreseen to occur – in such a way that they will not only be seen as search engines, but rather as tools for retrieval, searching and summarizing, and employed in (virtual) environments that are familiar to the users and that are adapted to their preferences. Being active in our 'own' hypermedia environments has become part of our daily activities, during our work, at home and during times of recreation. WWW and its tools have

☆ GOOGLE is a trademark of Google Inc.

E-mail address: anijholt@cs.utwente.nl

0020-0255/$ - see front matter
doi:10.1016/j.ins.2007.08.026

become part of our personal environment, whether this environment is embedded in our home, office or mobile environment. Through bookmarks, self-created web pages with interesting links, audio and web camera links that provide access to environments inhabited by friends, colleagues or others that we want to relate to, we can build our own personalized and real-life web environment and tools to do so will be provided by companies that are now associated with browsers and search engines. The environments allow us access not only to previously stored and transformed information, but also access to real, mixed reality and virtual reality worlds, to access and take part, in real-time, in events and to act as a member of self-chosen communities in which we can also display interest, self-disclosure and become involved in shared activities with family, relatives, friends and colleagues.

Obviously, it looks like a long way to go, from current search engines to personalized engines and personalized searchable and otherwise accessible environments, where the environments allow querying and attending real-time events and querying and re-experiencing previously stored events. However, presently we see three developments taking place in parallel.

(1) Companies such as Microsoft, Google, Yahoo and others are continuously extending their services to their users (chat environments, blogs, domain search engines, natural language access, picture search, sharing files, sharing music, sharing photographs, 2D maps and satellite images, annotated 3D Earth maps, etc.).
(2) Ambient intelligence (AmI) research has become a leading paradigm in human–computer interaction and ubiquitous computing. Inhabitants of AmI environments expect social and intelligent support from the environments they visit or live in. The environment is attentive and pro-active and supports multiparty interaction.
(3) 'Electronic Chronicles', 'Memories for Life', and 'Lifelogs' are among the terms that are used to denote the research area dealing with the capture, analysis, interpretation and storage of temporal streams of data. Surveillance data is one example, but also data obtained from wearable and mobile sensors, patient monitoring, biometrical information, video recordings, etc.

In this paper, we discuss these developments and show how they will be integrated in future (virtual) home, office and public environments. We will look at technology that is being developed in European research projects on multi-modal and multi-party interaction in order to support our views on the integration of these three developments in the near future. In particular, we will explore current and future ambient intelligence research and technology, and look at the way results from research and development done in the context of some research projects on the design and development of meeting support technology (smart meeting rooms, remote meeting participation, distributed meetings, distributed collaborative work spaces, etc.) can be explained and explored in the context of the ambient intelligence point of view on home environments and in the context of an electronic chronicles point of view of collecting and exploring personal data. One recurring point of view will be the ability to store human activity, to take part in (i.e., experience) human activity, to support human activity (based on some level of understanding of the human activity), and to re-experience activities. This re-experience may go from looking at the electronic minutes of a meeting to re-experience, in an immersive way, your own wedding.

1.1. Contents of this paper

In Section 2 of this paper we have some global observations on searching, browsing and visiting virtual environments (not necessarily virtual reality environments) and developments that allow users to design personal environments and environments users want to share with others (family, friends, colleagues, and everyone).

Section 3 is devoted to ambient intelligence technology and environments. Ambient intelligence (AmI) is about ubiquitous computing and social and intelligent interfaces [49]. Inhabitants and visitors of AmI environments obtain support from these environments in their activities, including their design of personal and shared environments as discussed in Section 2. Their personal and shared environments have become AmI environments that allow searching, browsing and visiting. In AmI environments, we have sensors (including cameras and microphones) that capture events and activities. In order to provide support to humans involved in these events and activities they need to be interpreted, requiring theory and models about human behavior and social interaction, in particular, when these environments are inhabited by several agents (human, robotic or virtual), multi-party interaction. Remote (on-line) access to AmI environments and off-line access to stored events and activities are issues that are discussed and that are part of research on capturing, representing, organizing, analyzing, and presentation of temporal streams of data [14]. Ultimately, rather than talking about access to stored information about events and activities, we prefer to talk about re-experiencing events and activities. Clearly, virtual reality environments are then the most obvious ways of replaying or regeneration of events and activities in which users are involved. Nevertheless, providing someone asking a question about events in the past with a verbally expressed answer, a verbal summary, or a multi-media summary composed from the fission of several multi-media streams can be useful as well and do not require 3D virtual reality replay or advanced video manipulation. Whatever way of answering, summarizing or regeneration and replay is chosen, it need to be based on theory and computational models that allow interpretation of the captured events and activities.

Although much research is done in the context of the AmI paradigm, hardly any research on human activities in smart environments takes into account both real-time support, remote real-time participation, storing events and (multi-party) activities and off-line access and replay or re-experience of events and activities. Research on smart meeting rooms is an exception. In Section 4, we explain and review these research approaches from points of view that allow exportation to other research and application areas. Obviously, our views are very much biased by our own involvement in the European augmented multi-party interaction (AMI) project on multi-modal interaction modeling in meeting contexts. Among others, we present our case study on (distributed) virtual reality representations of meeting events and activities. Visualization, virtual reality, multi-modal interaction and embodied agents (virtual humans) play important roles. With appropriately equipped smart home environments we can as well support: (1) multi-party interaction and joint activities of family members (including robots, virtual pets and virtual humans), (2) real-time monitoring 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. Finally, Section 5 contains conclusions and has observations about future research.

2. Browsing, sharing, visiting, inhabiting, participating

In current commercial web environments and web tools there are ways for users to create their own environments and there are ways for the environments to adapt to their users. Concerning the latter, navigation behavior can be observed, patterns of behavior can be distinguished and depending on these patterns a browser can suggest navigation acts to a user in order to reduce inefficiency and the feeling of being lost in hyperspace. Search engines have been provided with natural language interfaces, it has become possible to retrieve pictures, and query results can be categorized according to relevance and to categories that fit the user's interests. Text, audio, pictures, video and virtual reality have become available on the web and can be queried. Navigation tools and web engines allow us to search and browse previously stored and automatically annotated (indexed) multi-media information. Recommendations to continue, to visit related websites or to buy products (articles for example) can be added. User profiles, domain dependency (see e.g., Google Scholar), context awareness and context histories are research issues from which we can expect a transformation from being a visitor to being an inhabitant of self-designed multi-media 'web' environments.

However, rather than searching and browsing multi-media that has been put on the web by people unknown to us, tools become available, also provided by browser and search engine companies, that allow non-professionals to design not only their own homepages, but also to share their diaries, their photo albums, their music and their video collections. It may be expected that tools will be developed and offered to the casual user to put more of her life on the web, share it with friends, relatives and possibly unknowns (as many people prefer to do) and that tools will be developed and offered for manual, semi-automatic and automatic annotation of this material in order to make it suitable for intelligent retrieval and browsing and for more advanced querying, including asking for a multi-media presentation of a selection of stored personal information. MyLifeBits [18] is an ambitious attempt of Microsoft Research to develop such tools. Tools that allow continuous archival of personal information, continuous registration of experiences and replay of experiences (for example, a trip replay visualization) are part of the project and this issue of replay will be discussed later in this paper.

Extrapolating current developments we assume that more and more personal information will be put on the web, but in such a form that the medium reflects the contents. There will be links between the various displays of personal information and the design of the personal information space will be done in such a way that designer and visitors (family, relatives, and friends) feel at ease in this environment since visual and audio clues are present that can be recognized and interpreted since they reflect knowledge that is available about the 'designer', the physical world and the community he or she is living in. An address of your home is not that interesting. A photo of your home is more interesting, a photo of your study provides even more information (preferably in the context of your home) and your webcam brings you even closer to those who are interested in you and the information you provide.

On the web, we can find virtual communities in which we can participate and can represent ourselves as avatars in virtual 3D environments. 'There' (http://www.there.com) and ActiveWorlds (http://www.activeworlds.com) are good examples of 3D chat and shopping environments. In ActiveWorlds you can build your own home and have it visited by other members of the community. While these artificial worlds allow the display of personal information through chat, choice of avatars and the design of buildings and rooms, there is hardly a sense of reality which makes it possible to feel at home. As is well known, many digital and 3D virtual cities are being developed. Many techniques have been developed, including satellite imagery and airborne laser scanning, to obtain 2D and 3D manipulative digital representations. 3D virtual cities, whether they are obtained by manual design or by automatic means, can be digital equivalents of real cities and visitors of these cities are provided 'with a genuine sense of walking around an urban place' [12]. These "true" virtual cities can be used to allow tourists, visitors or inhabitants to explore a particular city and to get a drive-through experience.

In this context, we should also mention Google's initiative to provide web users with Google Map and Google Earth. They allow interactive access to maps and satellite photos and, although presently only for a limited number of locations, 3D views of parts of cities. In addition to services provided by Google (finding a business, get directions, and obtain sightseeing information), users can build their own Google Map based services. That is, Google Map can be annotated with multi-media information, allowing traffic information, dating, and housing services. Millions of users will be able to add information and services, turning these environments into global property-based and more local, geographically based communities. In addition to access to text-based environments, future search and browse engines need to be able to provide access to these multimedia-annotated 3D environments. Apart from many 'serious' applications one can think of designing games and of providing other types of entertainment in these environments.

As a next step we may consider modeling our home environments in 3D virtual reality. Clearly, no aerial or satellite photos will help us to obtain such 3D representations. However, in ambient intelligent research there is already the assumption that future smart home environments are equipped with cameras, microphones and other sensors, and from the information that can be captured we cannot only build a virtual reality representation of the home environment, but we can also consider real-time support to activities performed in these environments and we can allow real-time remote access. On- and off-line searching, browsing and participating in such environments are issues that will be discussed in the next sections of this paper. In addition to Google Earth, and possible 'Google City' and 'Google Street', these possibilities make it acceptable to think of a Google Home that allows us to explore our home, our home activities and also allows us remote access to home activities and off-line access to our captured home activities and events.

In general, visualization and digitalization of real-world events in 3D virtual reality allows (immersive) browsing, searching and retrieval of events. Hence, it allows someone to become an observer of events that took place. There is no need to confine ourselves to a straightforward mapping of real-world events to events in a "true" virtual environment. Events can be given different representations (and visualizations) depending on a particular application or user, and extra information (obtained from context, user interests and history) can be added.

What becomes possible if we can do this in real-time? First of all, we can observe events taking place in reality in a virtual reality representation that allows us to take different viewpoints, including the viewpoints of the actors in the events, and it allows us to view and access meta-information that is provided by the built-in intelligence of the generated environment. Secondly, real-time generation allows real-time interaction with human and virtual agents in these environments and participation in joint activities and events. As mentioned, we will explore these possibilities in the next sections.

3. Ambient intelligence technology and environments

Environments equipped with Ambient intelligence technology provide social and intelligent support to their inhabitants. The majority of ambient intelligence research is on providing support to individuals living or working in these smart environments. However, in home and office environments we have also people interacting with each other and interacting with smart objects (e.g., a mobile robot, furniture, intelligent devices, and virtual humans on ambient displays). Cameras, microphones and other sensors can be used to detect and capture such activities. Obviously, this 'multi-party interaction' needs support by the environment as well, requiring theories and (computational) models of social and professional interaction.

In this section, we discuss support to individuals and parties that visit or inhabit social and intelligent home environments. We discuss remote participation in events that take place in such environments and we discuss off-line access to the captured information in social and intelligent home environments. This off-line access should also allow the replay of experiences. Clearly, some of these views are unusual. Therefore, we first look at our meeting paradigm. This paradigm will make clear why we extend the usual viewpoint on ambient intelligence, that is, to provide real-time support to activities taking place in a smart environment, with facilities to have real-time access to these activities, to memorize these activities, and to manipulate and replay these activities.

3.1. The meeting paradigm

There is one important domain of application of ambient intelligence technology where many of the viewpoints we mentioned above come together in a natural way. This is the domain of meetings supported by smart environment technology. In this domain it is useful to provide support during the meeting, it is useful to allow people who cannot 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 multi-media 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. Meetings involve multi-party interaction and looking at smart environments from the point of view of supporting multi-party interaction adds some interesting research issues to the area of ambient intelligence research.

Firstly, in order to be able to provide support, the environment is asked to understand the interactions between its inhabitants and between inhabitants and the environment or smart and may be mobile objects available in the environment. Although we see the development of theories of interaction and behavior, these theories are rather poor from a computational point of view and therefore they hardly contribute to the design of tools and environments that support activities of human inhabitants. Hence, the need for computational theories of behavior and interactions needs to be emphasized. Input can be obtained from sensors for sound, image, and haptics. The interaction that has to be perceived does not only include all aspects of focused interaction, but also aspects of unfocused interaction. Interpretation requires the fusion of all modalities that can be perceived by the environment into various levels of annotation schemes and semantic/pragmatic representations that allow further processing. Based on the interpretation and the resulting representation(s) the environment, its virtual inhabitants and its smart objects need to provide real-time support to the human inhabitants or visitors of the environment. They need to decide how to present this support, through which modalities, and with which content. On the one hand, there can be implicit and explicit calls for support by the inhabitant or visitor of the environment, on the other hand, the environment can decide that this particular person or group of persons can benefit from its previously obtained knowledge and may suggest or perform, preferably welcome, spontaneous real-time support.

A second research issue that needs to be mentioned is the real-time monitoring of activities, the on-line access to information about activities taking place and also the on-line remote participation in activities or influencing activities in smart environments. Clearly, topics associated with this issue are also present when we look at surveillance technology and computer-supported collaborative work (future workspaces).

The third research issue concerns the off-line access to stored information about activities in smart environments. This latter issue may involve retrieval, summarization, and browsing of raw data but also the display of newly composed multi-media presentations of the captured data. Automatic annotation of information coming from different input sources and fusion of information coming from different input modalities into a representation that allows support to the inhabitant or visitor of an environment also allows indexing and retrieval of events (hypermedia), browsing of activities, reporting and summarization, and a replay, e.g. in virtual reality, of what has been going on in a particular period of time or before, during and after a particularly interesting event in the environment.

Finally, controlling the environment and its inhabitants is an other issue. Capturing events into representations that allow retrieval, browsing, summarization and multi-media generation also allows others (owners, providers, visitors) to use this information to influence and control the inhabitants and visitors of these environments. Clearly, this issue is very much related to privacy questions, that is, who has access to this information and who owns the ambient intelligence environment? The inhabitants of an environment are spied on. How does this influence their behavior? Knowing that there are eyes and ears that observe their behavior in possibly unknown ways may have a negative impact on the natural behavior of inhabitants and visitors of ambient intelligence environments and therefore will have negative consequences for the performance of the environments. Due to these eyes and ears, available in natural objects and more or less hidden in the environment, we may even ask whether being the sole inhabitant of such an environment is in fact impossible. Being there assumes being part of a gathering and also assumes behaving as being in a public environment, including feelings of presence, co-presence, focused and unfocussed interaction behavior [19].

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. Sometimes we are only interested in providing real-time support to an individual entering an ambient intelligence environment. Sometimes we just want to monitor what is happening and having an alert when something unusual is going on. Sometimes we just want to know what has been going on while we were not present.

The points of view expressed in this section have emerged in the context of some projects on smart meeting environments. The models and technology developed in these projects will be discussed in Section 4.

3.2. *Social and intelligent home environments: support and looking back*

Whatever kind of situation we are in, when 'ambient intelligence' in one or other way is able to support our activities we can be happy with it. Maybe the activities can be done more efficiently due to this support or they can become more enjoyable. Do we want to look back at activities, do we want to retrieve information about previous activities or do we want to experience these activities again, may be from an other view point or being in an other's person skin?

Our viewpoint is that there are lots of reasons for wanting to look back on a previous activity in which we or our friends and relatives were involved. Spontaneous gatherings at home, family gatherings and, generally, meetings and joint activities with friends, relatives and family members differ from meetings. Meetings are structured and certain goals are defined in advance. Although meetings differ from joint activities in a home environment, also in home environments meeting support technology that is now developed in some large European projects can play useful roles. The home environment can ask for real-time support for activities that take place, sometimes it can be useful or enjoyable to remotely take part in home activities and sometimes we would like to experience in some or other way an important moment again. Presently this is done with diaries, photo albums and video collections. Web providers made it already possible to share these collections with others. Personal archives are made accessible for others and personal notes and thoughts appear in blogs on the web. This can be considered as a first step to a continuous registration of events in social environments

[11] and at the same time to technology that makes it possible to search, browse and replay such information or allow to get immersed in this information (see also [50]).

Currently, most ambient intelligence technology that is being developed concerns applications as home environment control and automation. Personal entertainment, health care and security are other application areas. In our view, we should also look at events that involve multi-party interaction for which real-time support is useful and where support requires some high-level interpretation (in contrast with turning on the lights when someone enters the room). This interpretation allows also for off-line intelligent search in the stored information, the development of intelligent browsing tools and multi-media presentation of the information. Among the possibilities for multi-media presentation we include ways of replaying, probably in a transformed and manipulated way of home activities (family meetings, visits of relatives, playing with children, a birthday party, a wedding, just an evening at home with everyone doing usual things, preparing a dinner in the kitchen, etc.).

Here, we will not go into details of Ambient intelligence research. Results of this research can be found in, among others, the yearly proceedings of the European Symposia on Ambient intelligence [1,28,4], and in many conferences on ubiquitous and pervasive computing. Neither do we have detailed looks at attempts for the continuous archival and retrieval of personal experiences. We refer the reader to the Memories of Life approach [14], DARPA's, LifeLog project (http://www.darpa.mil/ipto/Programs/lifelog/) and Microsoft's MyLifeBits [18,3]. It may be useful, when capturing information, to make a distinction between information obtained from body area networks, personal area networks, local area networks, wide area networks and Cyberworlds [42]. Having a (mixed reality) 'album' of important (personal) events is one of the streams (*My Life Album*) of the *IntoMyWorld* candidate Presence II project [48]. Among the examples that are mentioned is the possibility to allow people to re-immerse themselves in their own weddings. Storing and replaying experiences is discussed in [8,9]. We will return to the latter in the next subsections and in Section 4.

3.3. The role of autonomous and semi-autonomous embodied agents

Ambient intelligence research requires user modeling. The environment and its smart objects and devices need to be aware of the characteristics and the preferences of the user. And, since the environment may have multiple interacting users it needs to be aware of relationships between these users. Users can be modeled as agents and, when we invite virtual visitors from remote places to our environment (family members that are away, relatives that are not able to join a party, colleagues or partners in an entertainment game), then it can be useful to have them and ourselves represented as 3D embodied agents. These agents are real-time controlled by the behavior of their human equivalents. The mapping from human behavior to embodied agent behavior does not have to be one-to-one. Their may be changes of appearance, movements or cognitive and perceptual abilities. There may even be some autonomy added to these agents. This autonomy may make it possible that even when its human equivalent is not alert, not present or not interested, it nevertheless can take some actions. An agent can change from semi-autonomous behavior to human-guided and human-controlled behavior.

More fully autonomous agents can be present as well in an ambient intelligence environment. Obviously, any Ambient intelligence environment has reactive and pro-active agents that somehow 'control' the environment in accordance with the preferences, the concrete actions and the global behavior of its inhabitants. For some of the roles played by these agents it seems useful to add embodiment to the features of these agents. That is, in addition to the synthetically embodied equivalents of human beings that may inhabit the environment, we can have virtual humans inhabiting the environment playing roles that act as an intermediate between the cognitive and social intelligence embedded in the environments and its inhabitants and visitors. When we introduce this embodiment we should be aware of the effects of the "Media Equation" [38] and we should try to turn the effects of this viewpoint into an advantage in designing useful relationships between embodied agents and their human partners.

In this well known book the authors report about experiments on humans assigning human characteristics to computers. It became known as the "social reactions to communication technology" perspective in which "computers are social actors". Made clear by experiments, it is not only a matter of contributing personality characteristics to computers; it is also a matter of being influenced by these properties while communicating. The book's conclusion?

> "Our strategy for learning about media was to go to the social science section of the library, find theories and experiments about human-human interaction - and then borrow. We did the same for information about how people respond to the natural environment, borrowing freely. Take out a pen, cross out "human" or "environment," and substitute media. When we did this, all of the predictions and experiments led to the media equation: People's responses to media are fundamental social and natural."

In various experiments the findings of Reeves and Nass have been refined and confirmed. Not only for 'media' in general, but also for the embodied agents which we do want to play roles in Ambient intelligence environments [16,10]. For a critical look at the Media Equation, advising more refined approaches, the reader is advised to consult [44].

Remarkably, looking at the experiments underlying the research presented in this book and looking at the experiments designed after the publication of this book, the so-called 'natural environment' does not really play a role in the observations in the book and the experiments that were designed. That is, rather than to rely on these authors' observations, in future research we have to look at the interaction characteristics of human–environment interaction and design our own research.

We should mention that it is not unusual to contribute personality characteristics to a room, a house, a mall, a street or square, to a town or even to a landscape or another natural environment. On the one hand, one may think that thoughts and activities (i.e., interactions with the environment) are influenced by the particular environment, on the other hand, users or inhabitants may choose a particular environment, may adapt the environment to their preferences and, whatever they do, leave their traces and because of that, their personalities in these environments. There are links between individuals and the physical environments they occupy [20]. Similarly, we may assume that whenever technology allows, consciously and unconsciously, links are created between individuals and their (Ambient intelligence) environments.

A next issue that needs to be put on the Media Equation research agenda is being part of a community of agents. Until now the Media Equation has been looked at from the point of view of the computer as a social actor and the point of view of embodied agents or talking faces as social actors. Above we mentioned a possible view that has been mentioned but not investigated in which the environment is considered as a social actor. In the context of Ambient intelligence we need as well investigate how humans interact within a community of embodied agents that is situated in a home environment.

The observations above should make us aware of what users will think and what they will expect when they communicate with embodied agents in their personal, their mobile and their work environments. Obviously, we can learn form situations where embodied agents are already employed. They have found their way already on commercial websites, in museum and other cultural heritage environments, and in educational, training and entertainment environments. Companies have been founded that design embodied agents on demand for such applications. Hence, there are already lots of embodied agents available, but they have extremely limited cognitive and social intelligence. The more remarkable, although being in accordance with the views expressed in the "Media Equation", in situations where these agents have been employed we see a more than significant impact on the behavior of the humans they interact with. When communicating with an embodied agent, humans reveal much more personal information, have a more careful language use and accept much more suggestions and recommendations [27].

When looking at Ambient intelligence home applications we can easily think of a long list of useful home and personal agents that can make our life easier, more interesting, more secure and healthier. In fact, the home inhabitant should have the means to introduce agents that suit his or her interests and needs. Clearly, other (embodied) agents (relatives, friend, and colleagues) can be allowed to have access to our home environment and may have the ability to communicate with our virtual agents, but also to physical robot agents that move around in our environment.

Obvious embodied agents we would like to see perform in our environments are agents that allow interaction about home control and automation, agents that are responsible for our agenda, agents that know about our history, our family, relatives, friends and colleagues, agents that help us to prepare meals, health or fitness agents, agents we can play games with, agents that mediate between us and companies that supply us with goods and services, and agents that we can consider as our personal assistant, butler or friend. As examples we want to mention, Maior-Domo [17], Laura [6] and the virtual room inhabitant [26]. Maior-Domo is a

Fig. 1. Maior-Domo helps the user performing tasks in the kitchen and living room.

domotic controller represented as avatar (see Fig. 1) that acts in a home lab situation where a real kitchen and living room have been built. This embodied agent helps the user to prepare a meal, to create a shopping list, program the washing machine and with other domestic tasks. The user is wearing a wireless microphone to have her conversation with the embodied agent. Laura acts as a personal trainer who advises you, depending on what you tell her about your recent activities and mood, about physical exercises. Facial expressions, body movements and gestures help to make Laura believable and her advice acceptable. Laura is a typical example of an agent that acts better when there is the possibility to develop a personal relationship between agent and human partner [33,46]. The Virtual room inhabitant knows about the human inhabitant and therefore is able to offer situated assistance. Interesting is that this virtual agent is able to move along the walls, the smart devices and smart objects that are available in the environment.

3.4. Coping with embodied agents: some design considerations

Without making any distinction, for the moment, between the different types of embodied agents that can usefully inhabit a home environment, it is clear that we need models of multi-party interaction (cf. [47]) rather than models of traditional human–human or human–computer interaction. Being able to model the external display of verbal and nonverbal interactions using interaction acts, interaction history, and interaction representation theory, requires, at a deeper level, the modeling of the beliefs, desires and the intentions of the individual task-oriented agents. Beliefs are about what the agent knows, desires are long-term goals and intentions are about the next steps the agent intends to take, taking into account its long-term goals, the contextual constraints and its capability to reason and to plan. Apart from contextual constraints that guide the agent's reasoning and behavior, there are constraints on behavior that follow from general models that describe emotions (emerging from an appraisal of events, from the point of view of goals that are pursued, taking place in the environment). A model of emotion synthesis that has become the standard (event appraisal) model for emotion synthesis is the so-called OCC model [36]. Among the appraisal variables are desirability, urgency or unexpectedness. Causal attribution is another issue (who should be blamed or credited) and so is the coping potential. A coping response can be problem-focused (where the agent decides to act on the world) or emotion-focused (where the agent decides to change its beliefs). In this way not being able to reach a certain goal may also have impact on the existing beliefs and desires of an agent. When an agent realizes that it cannot reach its goals, it can decide to cope with its emotions of disappointment by adapting its beliefs and goals. Both appraisal and coping need to be modeled [21]. In current research it is also not unusual to incorporate a personality model in an agent to adapt the appraisal, the reasoning, the behavior, and the display of emotions to personality characteristics. A well-known personality model that is often used in agent design is the five-factor personality model based on five personality dimensions (openness, conscientiousness, extraversion, agreeableness, and neuroticism) [31].

Clearly, agents involved in multi-party interaction not only have goals that follow from short-term and individual benefits that can be reached, but they can also take into account goals that are pursued by a community of agents and they can also take into account social relationships that exist between agents. As mentioned before, when we talk about agents, these agents can be humans taking part in the interaction, virtual humans (autonomous agents) that take part in the interaction and embodied agents that represent humans that take part in the interactions. When we talk about goals of a community of agents, we need to talk about cooperation between agents and how social relationships influence cooperation. Clearly, agents can be designed to be responsible, helpful and cooperative. While acting in a virtual environment they can take into consideration their own benefits, the benefits of society or the benefits of both themselves and the society. It means that they need to get involved in social decision-making [22] and they need to be aware of the effects of their acts with respect to themselves and their society. In these situations an agent needs other agents to achieve its intended goal and so social dependencies become important. An agent can have social power over other agents [7].

Finally, when we put humans and embodied agents in shared environments we should take into account the question why they share a particular environment and how we can make use of that kind of knowledge in order to obtain a better interpretation of what is or has been going on in an environment. Does the environment aim at collaboration, entertainment, health improvement, home work, discussing the past? Understanding what is going on in a particular (mixed-reality) environment in order to allow real-time support and off-line access to captured information requires understanding of the tasks and the domain associated with the environment. This requires also, as argued above, going from all kinds of existing agent theories that start with beliefs, desires and intentions, to agent theories that try to take into account interaction subtleties, interaction rituals and emotions associated with interactions. For example, depending on the application, we need to look at theories of how people behave, in home situations and in public spaces [5,19].

It is certainly not our intention here to survey all existing agent theories that we expect to be useful in the context of Ambient intelligence home environments. However, from our observations it should be sufficiently clear that when we introduce human and virtual (embodied) agents in these environments, the above mentioned aspects have to be dealt with in order to understand and support social and intelligent interactions in the environments.

3.5. Capturing human activities in context

In Ambient intelligence environments human activities need to be captured in order for the environment to provide, after interpreting the activities, real-time support to humans involved in these activities. And, of course, allowing off-line access to the captured information. Access may mean being able to ask questions about specific events, questions about persons involved in these events and why they did behave the way they did, asking a summary of events or asking for a personalized answer, summarization, or replay of events that took place and that have been captured. There is a lot of technology available to capture events in a physical environment. However, to do this in an unobtrusive way is a problem and to do this in a way that allows the fusion of information coming from different streaming information sources and a subsequent interpretation is even more a problem. Nevertheless, examples that demonstrate this are available. See also Section 4.

Video-capturing of environments and transforming video images to 3D virtual reality representations of these environments is a well-established research area. Reconstruction of environments in virtual reality can be done in real-time. However, it is also possible to have the environments downloaded from a database. An interesting approach can be found in [45] where a PDA is introduced that can be used as interface to all devices in the ambient intelligence environment. When entering a room the 3D scene of room is loaded from a database and the available devices are discovered and positioned in the 3D scene. The PDA allows access to the devices through the 3D interface. Hence, here we have a real-time positioning of the devices that can be accessed in a virtual reality representation of an environment.

However, we need capturing of human activity and multi-party interaction in order to be able to provide real-time support. In order to replay or re-experience certain events and in order to retrieve events or ask questions about them, we also need to store events. That is, we need to be able to store multi-media information about events and we need to be able to present, transform and recompose multi-media information. One possible way to do this is to make use of virtual reality technology, that is, to regenerate events in a 3D virtual reality representation. If this can be done in real-time with the capturing of the events, it also becomes possible to provide real-time virtual access to activities and have participants that are geographically dispersed to share the same virtual environment.

4. Smart and distributed meeting environments

> "What do people do at work? They go to meetings. How do we deal with meetings? What is it about sitting face-to-face that we need to capture? We need software that makes it possible to hold a meeting with distributed participants – a meeting with interactivity and feeling, such that, in the future, people will prefer being telepresent."
> Bill Gates, 1999.

4.1. General background and introduction

By looking at the earlier mentioned AMI project we want to make clear that technology obtained in multi-party interaction research as is now becoming available, can be usefully employed in the context of other smart environments. The AMI[1] project builds on the earlier M4 project (Multi-Modal Meeting Manager). Both projects are concerned with the design of a demonstration system that enables structuring, browsing and querying of archives of automatically analyzed meetings. The meetings take place in a room equipped with multi-modal sensors. Multi-media information captured from microphones and cameras are translated into annotated multi-media meeting minutes that allow for retrieval, summarization and browsing. The result of the M4 project was an off-line meeting browser.

More than in M4, in the recently started AMI project attention is on multi-modal events. Apart from the verbal and nonverbal interaction between participants, many events take place that are relevant for the interaction and that therefore have impact on their communication content and form. For example, someone enters the room, someone distributes a paper, a person opens or closes the meeting, ends a discussion or asks for a vote, a participants asks or is invited to present ideas on the whiteboard, a data projector presentation is given with the help of laser pointing and later discussed, someone has to leave early and the order of the agenda is changed, etc. Participants make references in their utterances to what is happening, to presentations that have been shown, to behavior of other participants, etc. They look at each other, to the person they address, to others, to the chairman, to their notes and to the presentation on the screen, etc. Participants have facial expressions, gestures and body posture that support, emphasize or contradict their opinion, etc.

To study and collect multi-modal data smart meeting rooms are maintained by the different research partners. They are equipped with cameras, circular microphone arrays and, recently introduced, capture of whiteboard pen writing and drawing and note taking by participants on 'electronic paper'. Participants also have lapel microphones and cameras in front of them to capture facial expressions.

4.2. AMI: from signal processing to interpretation

The meeting support application researched in the AMI project [29] requires the development of tools that take into account the meeting context. Rather than zooming in on constraining general methods of detecting and interpreting events in physical environments, we have a bottom-up approach starting with observed events in meeting environments and attempting to model and explain them using more general observations on theories of verbal and nonverbal communication.

[1] AMI (http://www.amiproject.org/ started on 1 January 2004 and has duration of three years. It is supported by the EU 6th FP IST Programme (IST IP project FP6-506811). AMI is succeeded by yet an other three year project, called augmented multi-party interaction with distant access (AMIDA).

Models are needed for the integration of the multi-modal streams in order to be able to interpret events and interactions. These models include statistical models to integrate asynchronous multiple streams and semantic representation formalisms that allow reasoning and cross-modal reference resolution. Apart from the recognition of joint behavior, i.e., the recognition of group actions during a meeting, there is also the recognition of the actions of individuals, and the information fusion at a higher level for further recognition and interpretation of the interactions.

When looking at the actions of the individuals during a meeting several useful pieces of information can be collected. First of all, there can be person identification using face recognition. Current speaker recognition using multi-modal information (e.g., speech and gestures) and speaker tracking (e.g., while the speaker rises from his chair and walks to the whiteboard) are similar issues. Other, more detailed but nevertheless relevant meeting acts can be distinguished: for example, recognition of individual meeting actions by video sequence processing.

Presently models, annotation tools and mark-up languages are being developed in the project. They allow the description of the relevant issues during a meeting, including temporal aspects and including low-level fusion of media streams. In our part of the project we are interested in high-level fusion, where semantic/pragmatic (tuned to particular applications) knowledge is taken into account (see e.g., [32]), i.e., we try to explore different aspects of the interpretation point of view. We hope to integrate recent research in the area of traditional multi-modal dialogue modeling. These issues will become more and more important since models, methods and tools that need to be developed in order to make this possible can be used for other events taken place in smart and ambient intelligence environments as well.

4.3. Progress and research results

In this section we review in some more detail the research themes of the AMI project and we illustrate some of the themes with results that have been obtained. We will look at data recording and annotation, at meeting modeling, at audio–video processing, and at providing access to multi-modal meeting data. We end this section with a few observations on real-time support during meetings and meeting assistants.

4.3.1. Data recording and annotation

A large effort has been the collecting of the AMI Meeting Corpus consisting of 100 h of multi-modal meeting data [30]. The meetings are in English, often with non-native English speakers as meeting participants. The data allows empirical observations and the training of statistical models, for example, for speech recognition, for gesture and body pose recognition, the recognition of meeting activities and gaze and turn taking behavior of participants. Machine learning techniques are based on manually annotated meeting data. The techniques aim at developing techniques for automatic recognition of properties that have been annotated explicitly in the training sets. Obviously, the data is also analyzed with the aim to obtain models that allow the design of rules and algorithms for extracting properties of the meeting data. Recognizing such properties underlies the interpretation of meeting activities.

In the corpus there is approximately 65 h of scenario-driven meeting data and about 35 h of natural meeting data. The scenario-data has been elicited using a design task, the design of a new type of television remote control. The participants played different roles (project manager, marketing expert, user interface designer, and industrial designer). Four design phases have been distinguished (kick-off, functional design, conceptual design and detailed design) and for each of these phases meetings were organized.

The rooms in which the meetings were recorded were equipped with microphones, both for close-talking and far-field audio, and with cameras capturing close-ups of the participants and cameras that capture global room views. In Fig. 2 some sample camera views are shown.

Tools have been developed to annotate the meeting data that has been captured. In Fig. 3 we show a tool developed by us to annotate the emotions of the meeting participants as they are perceived by the annotators when listening to and looking at a particular meeting participant. In the left window the video from a close-up camera is shown, in the right window the annotator marks the emotions that are perceived. Emotion annotation is just one example of a type of annotation, and obviously, in order to extract useful information from a corpus, it need to be combined, as any other type of property annotation, with all kinds of annotations that

Fig. 2. Sample camera views in the Edinburgh meeting room.

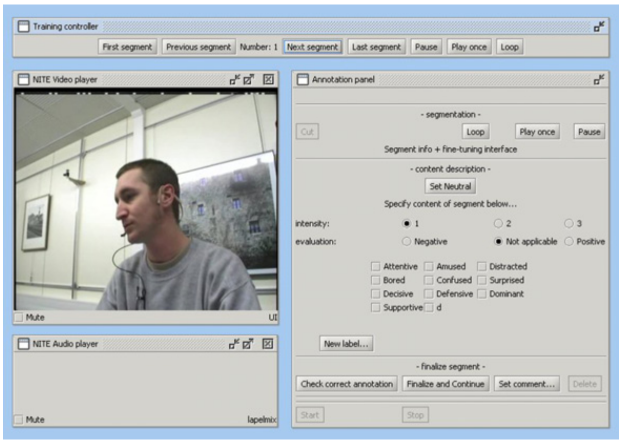

Fig. 3. The emotion annotation interface.

are considered to be useful for recognizing and interpreting meeting behavior and meeting events. That is, interdependencies of annotated phenomena need to be explored in order to allow us or an automatic extraction procedure to understand meeting activities. Among others, the following properties are annotated: speech transcription, location of individuals, dialogue acts, hand and head gestures, group activity, topic segmentation, emotion display, and focus of attention. The interdependencies of these properties, considered from the point of view of meetings, are input for possible models of meetings.

4.3.2. Meeting modeling

Having meeting models allows us to develop technology to give real-time support to meeting participants. These participants can be physically present in the same meeting room, we can have remote participants or we can have a situation where all meeting participants are distributed. Meeting models also allow us to structure and present meeting information in such a way that it can be more easily accessed, in an off-line manner, after a meeting, by both participants and others that are interested. Some objectives of meeting modeling that are in the core of the AMI interests are providing answers to questions such as "what is the current focus of attention for the group", developing turn-taking models for meetings (especially useful to provide real-time support in the case of distributed meetings, and the development of dialogue models that reveal the discourse structure. With these models querying and browsing of meeting data can be done more intelligently and, when the methods work in real-time, chairpersons and meeting assistants can use this information about the meeting to improve their performance and the meeting process.

Preliminary results have been obtained for evaluating successful meeting behavior. This has been done by using questionnaires and by considering the various input, process and output variables for meetings. More importantly for the subject of this paper, since in every ambient intelligence environment we need to consider multi-party interaction are results that have been obtained for modeling floor, turn taking and addressing behavior. Relevant features have been identified for the automatic recognition of addressees from visual focus of attention (gaze), speech and contextual dialogue parameters [23,24]. Argumentation modeling is another research issue in the context of meeting modeling. The ultimate goal is to design recognizers for argumentation episodes using multi-modal surface cues. A scheme for annotation of discussions and arguments has been defined and applied to 250 discussions in the AMI corpus. From the (presently manual) annotations argumentation diagrams can be obtained automatically. They are intended to aid the process of cognitive understanding of meeting discussions. The scheme and its rationale are discussed in [39].

4.3.3. Audio–video processing

Various recognition algorithms (among them HMMs, Bayesian networks and neural networks) have been ported to the AMI meeting domain and evaluated. Automatic recognition from audio, video, and combined audio–video streams is meant to provide us the means to: (1) recognize what is said by participants, (2) recognize what is done by participants (physical actions), (3) recognize where each participant is, at each time, (4) recognize participants' emotional states, (5) track what (person, object, or region) each participant is focusing on, and (6) recognize the identity of each participant. The main results have been obtained in the following areas [2].

Speech recognition. Verbal communication is the backbone of meetings. Automatic transcription of this communication is needed for meeting analysis, content analysis, browsing, retrieval and summarization. In addition to automatic speech recognition and system architecture issues there are issues such as speech activity detection, evaluation, keyword spotting and phoneme recognition (e.g., for speaker recognition).

Localization and tracking. Underlying many useful meeting recognition and interpretation tasks is the ability to detect and track multiple persons in the video sequences. Detecting and tracking of head, face and hands provides us with information about locations and it is a first step towards identifying people, face recognition, facial expression recognition and emotion recognition. Methods need to work under different real-world conditions and they have to deal with the problem of object and person occlusion. Statistical models and algorithms have been introduced for fusing multi-modal information obtained from different cameras and microphones and from multiple visual conditions.

Actions and gestures. Actions and gestures that occur frequently within meetings have been identified. The important ones are those that add to a semantic analysis of a meeting. Examples are certain head gestures (e.g., nodding and shaking), certain body gestures (e.g., leaning forward/backward, sitting down, standing up) and certain hand gestures (e.g., voting, pointing, writing). In addition there are speech supporting gestures (beats, iconic gestures, etc.). Algorithms for automatic extraction of features (using a model-based pose estimation program [37]) and automatic segmentation of feature streams have been designed and are evaluated.

4.3.4. Access to multi-modal meeting data

The multi-modal meeting data that has been captured needs to be accessed by users. This can be done off-line, for instance by meeting participants who want to verify what exactly has been decided or what they did promise, or by persons who could not make it to the meeting and need to know about decisions and the way they were made, the argumentation that was used [39] and who were in favor or against the final decision that was reached. In an on-line setting we can have a remote participant that participates in one or other way thanks to some kind of (visual) representation of meeting activities going on. To achieve these goals methods to structure and segment meetings are under investigation. They set the floor for information extraction, retrieval, summarization and browsing. The methods deal with syntactic chunking, dialog act classification and segmentation, topic segmentation, and meeting act recognition. Other research helpful for interpreting and retrieving meeting information deals with the earlier mentioned focus of attention research, addressee identification [23] and dominance detection [40].

For browsing meeting information a multi-media browser was developed (jFerret) in which research results can be embedded. It allows the presentation of video, audio, slides and annotation time-lines, but it also allows plug-ins that visualize the argumentation structure of a discussion or that show dominance levels of the meeting participants (see Fig. 4).

4.3.5. Real-time support

Clearly, since the research approaches and the research results mentioned in the previous subsections contribute to the design of smart (meeting) environments that understand what is going on in the environment, they can be used, assuming that they work in real-time, to assist meeting participants (distributed or not) dur-

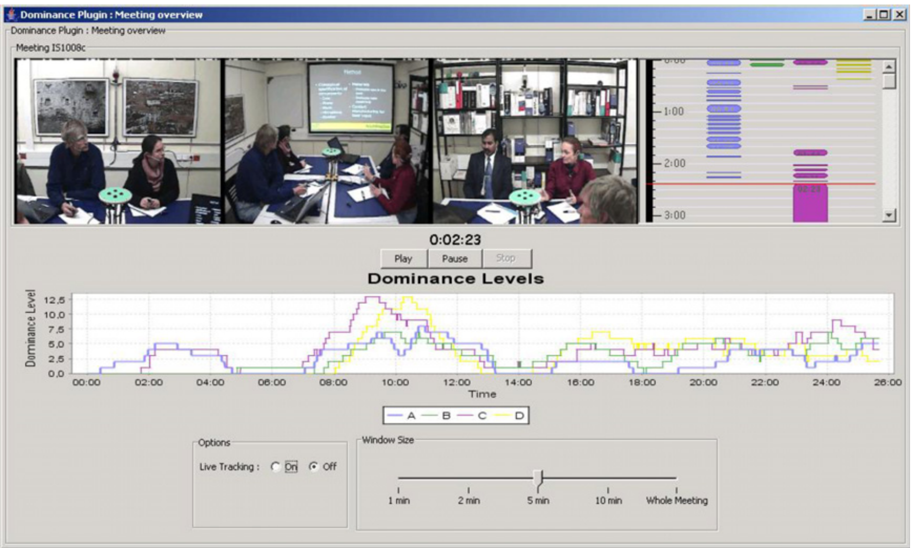

Fig. 4. Display of dominance information in a meeting browser.

ing their meeting activities. Meeting assistants that analyze activities and based on these analyses provide support to meeting participants become possible [41]. In the next section one particular approach to meeting visualization is explored.

4.4. Visualization, virtual reality representation and replay

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 studied posture and gesture activity, using our vision software package. A flock-of-birds package was used to track head orientation of some of our 4-party meetings. It allowed us to display animated representations of meeting participants in a (3D) virtual reality environment [34,35]. An early attempt to display meeting events in a virtual meeting room can be found in [15]. In an EU roadmap document of future workspaces we find similar ideas [13]. Presently, various kinds of 3D reconstruction technology allow the reconstruction of events in virtual reality environments. In our environment visualized events can be augmented with meta-observations provided by support agents and displayed in the virtual environment. This is illustrated in Fig. 5.

Even more attractive is to have meetings represented in a virtual meeting room (VMR), where participants do not share the same physical space. We introduced a prototype version of a distributed meeting room set-up. This set-up allows the connection of several inhabited smart meeting rooms and the representation of the participants and their activities in a shared virtual environment, made accessible for participants (and observers) in real-time. It allows the participants to take part in the meeting, perceiving the verbal and nonverbal communication by other participants through their avatars, from their assigned position around the meeting table. As shown in Fig. 6, also in this distributed version we can add meta-information about the meeting and its progress to the visualization of the virtual room.

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 [37], 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

Fig. 5. The virtual meeting room showing speaker, gestures, head movements, and addressee(s).

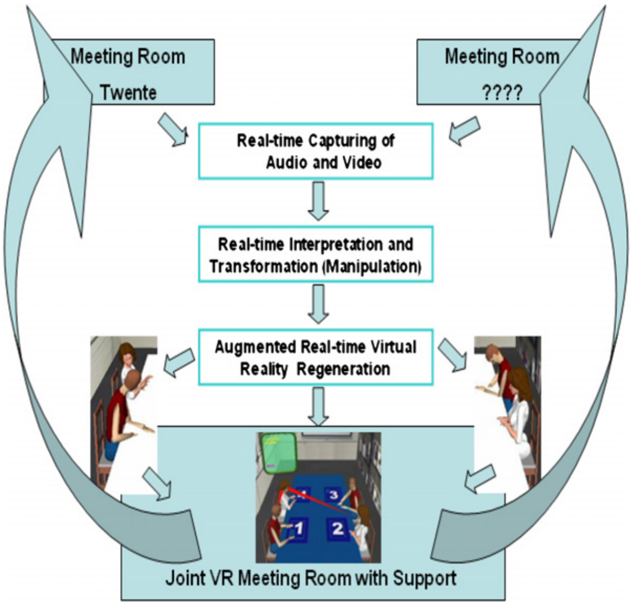

Fig. 6. Capturing, manipulation and re-generation of activities in remote locations in a joint virtual meeting room.

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 make 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, etc. 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.

5. Conclusions

In this paper we introduced some general observations about ambient intelligence in the home environment. Home automation is important, but providing real-time support to the inhabitants during their activities is important as well. This real-time support requires interpretation of home activities. In many of these activities we have to deal with multi-party interaction. That is, there are verbal and nonverbal interactions between the human inhabitants of the environment. Moreover, with the introduction of mobile robots, smart objects and virtual embodied agents displayed on walls and objects, the multi-party members will also include these artificial and pro-active agents. The environment needs some understanding of such interactions and therefore we need to look for models for multi-party verbal and nonverbal interactions.

Meetings are rather controlled events and therefore they are a more acceptable target for preliminary research in this direction. For that reason we looked at the approaches and the (preliminary) results that are obtained in the European augmented multi-party interaction (AMI) project on smart meeting environments. In this project, real-time support is only one of the objectives. Rather the emphasis is on querying and browsing the multi-media information that has been captured using various types of sensors. Being able to replay in one or other form of a meeting is also an interesting objective. May be not at first sight, but these additions to real-time support are useful in home environments as well. In fact, there are many examples of such research projects in the area of ambient intelligence. Moreover, rather than connecting distributed meeting participants or smart meeting rooms in a virtual reality meeting room as we discussed in Section 4, we can also consider virtual visits to a smart home environment or virtual participation in activities (e.g. a birthday party) in a home environment. Very often an interactive wall is the interface between spatially distributed participants. In the RemoteHome [43] there is not only a wall, but also furniture and other objects that can communicate with their counterparts in another room. Many other examples where physical objects act as interface, allowing social interaction, between remote spaces exist (see e.g. [25]). Yet another example is the hug suit, a wireless, sensor-rigged 'jacket', developed at the Nanyang Technological University in Singapore. Such a suit, worn by the child, allows parents that are away from home to use the internet to give their children a (virtual) hug. That is, the jacket transforms received signals in vibrations (pressure changes) and temperature. Instead of using keyboard and mouse it is foreseen that pets, for example a teddy bear, can be used to record the parental hugs and transmit them over the Internet Apart from real-time support to home inhabitants and real-time remote access from other smart environments, the meeting technology that has been discussed in Section 4 also allows intelligent querying, browsing and replay of previous interesting home events. Obviously, going 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 AMI and some other large EU projects we now see, as discussed in this paper, that small steps in this direction are taken.

Acknowledgements

This work was partly supported by the EU 6th FWP IST Integrated Project AMI (Augmented Multi-party Interaction, FP6-506811, publication AMI-168). We described joint work of the Human Media Interaction (HMI) group of the University of Twente and some work, particularly reported on in Section 4.3 of this paper, by our partners in the project.

References

[1] E. Aarts, R. Collier, E. van Loenen, B. de Ruyter (Eds.), First European Symposium on Ambient Intelligence, November 3–4, 2003, Lecture Notes in Computer Science, vol. 2875, EUSAI, Veldhoven, The Netherlands, 2003.

[2] M. Al-Hames et al., Audio–Visual processing in meetings: seven questions and some AMI answers, in: S. Renals, S. Bengio, J. Fiscus (Eds.), Proceedings Machine Learning & Multimodal Interaction (MLMI 2006), Lecture Notes in Computer Science vol. 4299, S. Renals, Springer, Berlin, pp. 24–35.

[3] A. Aris, J. Gemmell, R. Lueder, Exploiting location and time for photo search and storytelling in MyLifeBits, MSR-TR-2004-102, October 2004.

[4] G. Bailly, J.L. Crowley (Eds.), Proceedings of the Smart Objects & Ambient Intelligence Conference, (SOC-EUSAI), Grenoble, France, October 12–14, 2005. <http://www.soc-eusai2005.org/proceedings/>.

[5] R.F. Bales, Social Interaction Systems. Theory and Measurement, Transaction Publishers, New Brunswick, 2001.

[6] T. Bickmore, Relational agents: effecting change through human–computer relationships, MIT Ph.D. Thesis, February 2003.

[7] R. Conte, C. Castelfranchi, Simulating multi-agent interdependencies: a two-way approach to the micro-macro link, in: K.G. Troitzsch, U. Mueller, N. Gilbert, J.E. Doran (Eds.), Social Science Microsimulation, Springer, Berlin, 1996.

[8] N. Correia, L. Alves, J. Santiago, L. Romero, Storing and replaying experiences in mixed environments using hypermedia, in: Memory and Sharing of Experiences Workshop, Pervasive 2004 Conference, Vienna, Austria, April 2004.

[9] N. Correia, L. Alves, R. Sá, J. Santiago, L. Romero, HyperMem: a system to store and replay experiences in mixed reality worlds, in: 2005 International Conference on Cyberworlds, November 23–25, Singapore, pp. 83–90.

[10] K. Dautenhahn, Socially intelligent agents in human primate culture, in: S. Payr, R. Trappl (Eds.), Chapter 3 in Agent Culture: Human–agent Interaction in a Multi-cultural World, Lawrence Erlbaum Associates, Publishers, London, pp. 45–71.

[11] M. Deutscher, P. Jeffrey, N. Siu, Information capture devices for social environments, in: Proceedings EUSAI 2004, Lecture Notes in Computer Science, vol. 3295, Springer-Verlag, Berlin, Heidelberg, 2004, pp. 267–270.

[12] M. Dodge, A. Smith, S. Doyle, Urban Science. GIS Europe, October 1997, vol. 6, No. 10, 26–29.

[13] F. Fernando et al. (Eds.), Future Workspaces. A strategic roadmap for defining distributed engineering workspaces of the future. Final Roadmap, IST-2001-38346, Deliverable 5.4, May 2003.

[14] A. Fitzgibbon, E. Reiter, GC3: Memories for life: managing information over a human lifetime, pp. 13–16. <http://www.memoriesforlife.org/>.

[15] E. Frecon, A.A. Nou, Building distributed virtual environments to support collaborative work. VRST'98, Taipei, Taiwan, November 1998, pp. 105–113.

[16] B. Friedman (Ed.), Human Values and the Design of Computer Technology, Cambridge, CSLI Publications, Cambridge University Press, 1997.

[17] A. Gárate, N. Herrasti, A. López, GENIO: An Ambient Intelligence application in home automation and entertainment environment, in: Proceedings Joint soc-EUSAI Conference, Grenoble, October 2005.

[18] J. Gemmell, G. Bell, R. Lueder, S. Drucker, C. Wong, MyLifeBits: Fulfilling the Memex Vision. ACM Multimedia'02, December 1–6, 2002, Juan-les-Pins, France, pp. 235–238.

[19] E. Goffman, Behavior in Public Spaces. Notes on the Social Organization of Gatherings, The Free Press, NY, 1963.

[20] S.D. Gosling et al., A room with a cue: Personality judgments based on offices and bedrooms, Journal of Personality and Social Psychology 82 (3) (2002) 379–398.

[21] J. Gratch, S. Marsella, Tears and fears: Modeling emotions and emotional behaviors in synthetic agents, Proceedings of the Fifth International Conference on Autonomous Agents, Montreal, Canada, ACM Press, 2001, pp. 278–285.

[22] N.R. Jennings, S. Kalenka, Socially responsible decision making by autonomous agents, in: K. Korta, E. Sosa, X. Arrazola (Eds.), Cognition, Agency, and Rationality, Kluwer Academic Publishers, The Netherlands, 1999, pp. 135–149.

[23] N. Jovanovic, R. op den Akker, Towards automatic addressee identification in multi-party dialogues, in: Proceedings 5th SIGdial Workshop on Discourse and Dialogue, Cambridge, MA, USA.

[24] N. Jovanovic, R. op den Akker, A. Nijholt, A corpus for studying addressing behavior in multi-party dialogues, in: L. Dybkjaer, W. Minker (Eds.), Proceedings 6th SIGdial Workshop on Discourse and Dialogue, Lisbon, Portugal, 2005, pp. 107–116.

[25] K.G. Karahalios, K. Dobson, Chit chat club: bridging virtual and physical space for social interaction, in: Proceedings CHI 2005, Portland, OR, USA, 2005.

[26] M. Kruppa, L. Spassova, M. Schmitz, The virtual room inhabitant – intuitive interaction with intelligent environments, in: S. Zhang, R. Jarvis (Eds.), Artificial Intelligence 2005, Lecture Notes in Artificial Intelligence, vol. 3809, Springer-Verlag, Berlin Heidelberg, 2005, pp. 225–234.

[27] H. Maldonado, B. Hayes-Roth, Toward cross-cultural believability in character design, in: S. Payr, R. Trappl (Eds.), Agent Culture: Human–agent interaction in a multi-cultural world, Lawrence Erlbaum Associates, Publishers, London, 2004, pp. 177–196, Chapter 7.

[28] P. Markopoulos, B. Eggen, E. Aarts, J.L. Crowley (Eds.), Second European Symposium on Ambient Intelligence November 8–11, Lecture Notes in Computer Science, vol. 3295, EUSAI Eindhoven, Netherlands, 2004.

[29] I. McCowan, D. Gatica-Perez, S. Bengio, D. Moore, H. Bourlard, Towards computer understanding of human interactions, in: E. Aarts et al. (Eds.), Ambient Intelligence, LNCS 3361, Springer-Verlag Heidelberg, 2004, pp. 56–75.

[30] I. McCowan et al., The AMI meeting corpus, in: L.P.J.J. Noldus, F. Grieco, L.W.S. Loijens, P.H. Zimmerman (Eds.), Proceedings Measuring Behavior 2005, 5th International Conference on Methods and Techniques in Behavioral Research, Wageningen: Noldus Information Technology, 2005.

[31] R.R. McCrae, P.T. Costa, Toward a new generation of personality theories: theoretical contexts for the five-factor model, in: J.S. Wiggins (Ed.), The Five-factor Model of Personality: Theoretical Perspectives, Guilford, NY, pp. 51–87.

[32] A. Nijholt, Multimodality and ambient intelligence, in: W.F.J. Verhaegh, E.H.L. Aarts, J. Korst (Eds.), Algorithms in Ambient Intelligence, Kluwer, Boston, 2003.

[33] A. Nijholt, Where computers disappear, virtual humans appear, Computers and Graphics 28 (4) (2004) 465–476.
[34] A. Nijholt, J. Zwiers, J. Peciva, The distributed virtual meeting room exercise, in: A. Vinciarelli, J.-M. Odobez (Eds.), Proceedings ICMI 2005 Workshop on Multimodal Multiparty Meeting Processing, Trento, Italy, October 2005, pp. 93–99.
[35] A. Nijholt, Meetings in the virtuality continuum: send your avatar, in: T.L. Kunii, S.H. Soon, A. Sourin (Eds.), Proceedings 2005 International Conference on Cyberworlds, November 2005, IEEE Computer Society Press, Los Alamitos, USA, Singapore, 2005, pp. 75–82.
[36] A. Ortony, G.L. Clore, A. Collins, The Cognitive Structure of Emotions, Cambridge University Press, 1988.
[37] R. Poppe, D. Heylen, A. Nijholt, M. Poel, Towards real-time body pose estimation for presenters in meeting environments, in: V. Skala (Ed.), Proceedings of the 13th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision. Plzen, Czech Republic, 2005, pp. 41–44.
[38] B. Reeves, C. Nass, The Media Equation: How People Treat Computers, Televisions and New Media like Real People and Places, Cambridge University Press, 1996.
[39] R.J. Rienks, D. Heylen, E. van der Weijden, Argument diagramming of meeting conversations, in: A. Vinciarelli, J.-M. Odobez, (Eds.), Proceedings Multimodal Multiparty Meeting Processing, Workshop at the 7th International Conference on Multimodal Interfaces (ICMI), Trento, Italy, 2005, pp. 85–92.
[40] R.J. Rienks, D. Heylen, Automatic dominance detection in meetings using easily obtainable features, in: S. Renals, S. Bengio (Eds.), Revised Selected Papers of the 2nd Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms, Lecture Notes in Computer Science, vol. 3869, Springer, 2006, pp. 76–86.
[41] R. Rienks, A. Nijholt, D. Reidsma, Meetings and Meeting Support in Ambient Intelligence, in: A. Vasilakos, W. Pedrycz (Eds.), Ambient Intelligence Wireless Networking Ubiquitous Computing, Artech House, Norwood, MA, USA, 2006 (Chapter 18).
[42] G. Riva, P. Loreti, M. Lunghi, F. Vatalaro, F. Davide, Presence 2010: The emergence of ambient intelligence, in: G. Riva, F. Davide, W.A. Ijsselsteijnm (Eds.), Being There: Concepts, Effects and Measurements of User Presence in Synthetic Environments, IOS Press, Amsterdam, 2003, pp. 59–82.
[43] T. Schneidler, RemoteHome, 2004. The RemoteHome site: <www.remotehome.org>.
[44] N. Shechtman, L.M. Horowitz, Media inequality in conversation: how people behave differently when interacting with computers and people, in: SIGCHI-ACM CHI 2003: New Horizons, ACM, New York, 2003, pp. 281–288.
[45] A.A.N. Shirehjini, A novel interaction metaphor for personal environment control: Direct manipulation of physical environment based on 3d visualization, Computers Graphics 28 (2004) 667–675, Special Issue on Pervasive Computing and Ambient Intelligence.
[46] B. Stronks, A. Nijholt, P. van der Vet, D. Heylen, Designing for friendship: becoming friends with your ECA, in: A. Marriott, C. Pelachaud, T. Rist, Zs. Ruttkay (Eds.), Proceedings of Embodied Conversational Agents – Let's Specify and Evaluate Them! Bologna, Italy, pp. 91–97.
[47] D. Traum, J. Rickel, Embodied agents for multi-party dialogue in immersive virtual worlds, in: Proceedings of the 1st International Joint Conference on Autonomous Agents & Multi-Agent Systems, vol. 2, 2002, pp. 766–773.
[48] P. Turner, S. Turner, D. Tzovaras, Reliving VE day with schemata activation, in: M. Slater (Ed.), Proceedings 8th International Workshop on Presence: Presence 2005, London, 2005, pp. 33–38.
[49] A. Vasilakos, W. Pedrycz (Eds.), Ambient Intelligence, Wireless Networking, Ubiquitous Computing, Artech House, Norwood, MA, USA, 2006.
[50] S. Vemuri, W. Bender, Next-generation personal memory aids, BT Technology Journal 22 (4) (2004) 125–138.