# SMART EXPOSITION ROOMS: THE AMBIENT INTELLIGENCE VIEW[1]

Anton Nijholt, University of Twente
Centre of Telematics and Information Technology (CTIT)
PO Box 217, 7500 AE Enschede, the Netherlands
anijholt@cs.utwente.nl

**Abstract - We introduce our research on smart environments, in particular research on smart meeting rooms and investigate how research approaches here can be used in the context of smart museum environments. We distinguish the identification of domain knowledge, its use in sensory perception, its use in interpretation and modeling of events and acts in smart environments and we have some observations on off-line browsing and on-line remote participation in events in smart environments. It is argued that large-scale European research in the area of ambient intelligence will be an impetus to the research and development of smart galleries and museum spaces.**

## 1. Introduction

In documents of the European Commission we see the mentioning of "the real world being the interface". In particular the 'Ambient Intelligence' theme of the European 6th Framework Research Programme demands systems, which are capable of functioning within natural, unconstrained environments - within scenes. Notions of space, time and physical laws play a role and they are maybe more important than the immediate and conscious communication between a human and a display [7]. In a multi-sensory environment, supported with embedded computer technology, the environment can capture and interpret what the user is doing, maybe anticipating what the user is doing or wanting, and therefore the environment can be pro-active and re-active, just capturing what is going on for later use, or acting as an environment that assists the user in real-time or collaborates with the user in real-time. Ubiquitous computing technology will spread computing and communication power all around us. That is, in our daily work, our home environment and our recreation environments there will be computers with perceptual competence in order to profit from this technology.

In this paper we discuss examples of environmental interfaces, environments equipped with sensors that capture audio- and visual information, with the aim to see how we can translate research in home and collaborative work environments to museum and exposition environments. In particular we look at research done in the recently started 6th framework project AMI (Augmented Multi-party Interaction). In this project we are involved with research on the semantics of the interactions and the semantics of other events taking place in an environment. Events and interactions are of multimodal nature. Apart from the verbal and nonverbal interaction between inhabitants of an environment, many other events can take place that are relevant for understanding what is going on (people entering a room, looking for a chair, addressing a robot, walking to a particular object, etcetera). The environment

---

needs to be attentive, but it should also give feedback and be pro-active with respect to the visitors of the environment or the participants in a collaborative event in the environment.

Presently, models, annotation tools and mark-up languages are being developed. They allow the description of relevant issues including temporal aspects and low-level fusion of media streams. Corpora of annotated events will help to learn to anticipate certain interests of inhabitants and visitors and also to anticipate what will happen next in an environment. We will make some observations on translating the actions and sequences of actions by individuals in our domains to actions in the museum domain. Comparisons will be made with ideas reported in earlier situated interaction research in museum environments [3,4,9,10]. The important idea that should be made clear however, is that in previous decades artists and museum professionals certainly made use of augmented reality and media technology to furnish museums and galleries. However, due to the large-scale European interest, as shown in the 6th Framework Projects on Information Society Technologies, we can expect that now comes the time that instead of ad hoc use of this technology the scale and the resulting examples, the availability and the cost decrease will allow to use this ambient intelligence technology in museums as it will be used in other public buildings and home environments.

This introduction is followed by a section about the current AMI project. Section 3 zooms in on the methods, the models and the technology that is developed in these projects. Section 4 generalizes from the meeting domain to other domains, including museums and galleries.

## 2. AMI: A European Project on Multi-party Interaction

The AMI project builds on the earlier M4 project (Multi-Modal Meeting Manager). M4, funded by the EU in its 5th Framework Programme[2], is 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 multimodal sensors. The aim of the project is to design a meeting manager that is able to translate the information that is captured from microphones and cameras into annotated meeting minutes that allow for retrieval questions, summarization and browsing. In fact, it should be possible to generate everything that has been going on during a particular meeting from these annotated meeting minutes, for example, in a virtual meeting room, with virtual representations of the participants. The result of the project is an off-line meeting browser.

Clearly, we can look at the project as research on smart environments and on ambient intelligence. However, there is no explicit or active communication between user and environment. The user does not explicitly address the environment. The environment registers and interprets what's going on, but is not actively involved. The environment is attentive, but does not give feedback or is pro-active with respect of the users of the environment. Real-time participation of the environment requires not only attention and interpretation, but also intelligent feedback and pro-active behavior of the environment. It requires also presentation by the environment of multimedia information to the occupants of the environment.

---

[2]  M4 started on 1 March 2002 and has duration of three years. It is supported by the EU IST Programme (project IST-2001-34485) and is part of CPA-2: the Cross Programme Action on Multimodal and Multisensorial Dialogue Modes. AMI 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).

More than in M4, in the recently started AMI project attention is on multimodal 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 the 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.

In order to study and collect multimodal meeting data a smart meeting room is maintained by IDIAP in Martigny (Switzerland). It is 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 maybe in the future cameras in front of them to capture their facial expressions, rather than cameras for overviews. In Figure 1 we show a three-camera view of a meeting between four persons.

## 3. AMI: From Signal Processing to Interpretation

Models are needed for the integration of the multimodal 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. Presently there are two approaches that are followed. The first is the recognition of joint behavior, i.e., the recognition of group actions during a meeting. Examples are presentations, discussions and consensus. Probabilistic methods based on Hidden Markov Models (HMMs) are used [5]. The second approach is the recognition of the actions of individuals, and to fuse them 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 multimodal 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. Examples of actions that are distinguished are entering, leaving, rising, sitting, shaking head, nodding, voting



**Figure 1:** 3 cameras capturing a small meeting

**Figure 2:** Pointing, rising and voting

(raising hand) and pointing (see Figure 2). These are rather simple actions and clearly they need to be given an interpretation in the context of the global event. Or rather, these actions need to be interpreted as part of other actions and verbal and nonverbal interactions between participants.

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. [2]). I.e., we try to explore different aspects of the interpretation point of view. We hope to integrate recent research in the area oftraditional multimodal dialogue modeling [7]. 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. Real-time Support and Off-line Browsing of Acts and Events

Apart from M4 and AMI there are several other research projects concerned with the computational modeling of events that take part in smart environments. Closely related to AMI is for example the work done at the University of California, San Diego, which includes the development of methods for person identification, current speaker recognition, models for face orientation, semantic activity processing and graphical summarization of events. There is both work on intelligent meeting rooms as on smart environments in general (AVIARY: Audio-Video Interactive Appliances, Rooms and sYstems) [8]. The Ambiance project, done in the context of a European project, is also more general than 'just' an attempt to model meeting situations. Rather it looks at smart home environments [1], requiring much more modeling of the environment, including the many (smart and mobile) objects that can play a role in activities among inhabitants or between inhabitants and the global environment. In the museum domain we can mention several projects (and already existing environments) where there are explorations that involve equipping the visitor with handheld devices, motion tracking and wireless data transmission (see e.g. the HIPS project [3,4]). In this domain there is experience with interactive art, augmented reality and other media technology that has been designed for a particular artwork or exhibition by artist or museum professional. Methods, tools and technology developed in ambient intelligence research can however become part of the infrastructure of museums. The general structure we like to distinguish is the following:

- Understanding the domain, its inhabitants (visitors, participants, users), objectives and activities. E.g., in the meeting domain we distinguish between different kinds of meetings, objectives, groups and personalities; these features are responsible for different kinds of meeting strategies and behaviors of participants. Similarly, in the museum domain it is useful to distinguish visiting strategies. There exist classifications of museum visitors. In [10] characteristics for four categories are given: there is the grasshopper (hopping from one stop to the other, only a few stops during a visit, not following the designated routes), the ant (tries to be complete in his visit, takes his time, is studying the items in the exposition), the butterfly (not really sequential, selective, attracted to some items), and the fish (quick and superficial, glancing, no particular preferences). Features that help to classify include duration of a visit, the sequential or non-sequential behavior, the selectiveness, the number of stops, proximity to the exposition items, etc.). In MIT's Museum Wearable project a distinction is made between busy, greedy and selective visitors. In a semio-cognitive approach to museum consumption experiences Umiker-Sebeok [9] distinguished four strategies of reception where each strategy also defines a visitor's view on a exhibition: the pragmatic reception, where the gallery is seen as a type of (work)shop and emphasizing the utilitarian values; the critical reception, where the gallery is seen as museum and where non-existential values are emphasized (e.g., aesthetics of displays); the utopian reception, where the gallery is seen as an encounter session and where existential values are emphasized (e.g., what does it say about my relationships with others); and the diversionary reception strategy, where the gallery is seen as an amusement park and non-utilitarian values are emphasized.

- Uni- and multi-modal perception, recognition and interpretation of information coming from different sources (sensors), including audio, video, haptic and biometric information. Needed is annotation of this information (for off-line processing purposes) and alignment and fusion of this information on different levels of representation and for different levels of processing. There are many challenges for audio and video processing in smart environments. There are multiple sound sources, speech is conversational and there may be non-native speakers, to mention a few problems for speech recognition. For video processing we have to deal with unrestricted behavior of participants with variations of appearance and pose, different room conditions, occlusion, etc. Multi-modal syntactic and semantic information need to be extracted in order to recognize and interpret participant behavior, participant interaction and meeting events. For example, once the environment is able to map sensory data on different types of visitors, the next step is to anticipate, support and influence their behavior. This may include making suggestions that fit their behavior or drawing attention to items in the exposition that may interest the visitor but that require a different behavior.

- Interpretation for personalized support, generation and participation. For meetings it is quite natural to be able to retrieve information. That is exactly where minutes are made for. In the AMI project we allow different types of retrieval: straightforward questions (who was there, who said what, what was decided), but also questions about more global issues, asking for a summarization, discussions related to a certain topic and the replay of part of a meeting. An off-line meeting manager or intelligent meeting browser that has some understanding of the meeting events supports the retrieval and (re-)generation of this information. An on-line meeting manager would make it possible to support the meeting participants and would also facilitate, e.g. by alerting at points of interest or by guarding the turn-taking process, remote participants to take part. What can this mean for the museum domain? When we have visited an exposition we can bring our visit home, to browse through it or make it available to others. We can also allow remote visitors in real-time. In the HIPS project [3,4] visitors can book-

mark moments of their visit by pressing a button on their electronic handheld guide. These moments can include information about the position of the artwork, an image of the artwork or a personal comment. This allows the visitor to re-experience his visit, share it with others and help to plan a next visit. This is a very limited way of revisiting. We can as well design a revisit enhanced with multimedia presentations or in virtual reality with the information that has been collected during the real visit. There will be recognition during this revisit, but there can also be additions, depending on new information provided to the virtual visitor and taking into account a different viewing situation (at home, in a hotel room, et cetera). Rather than browsing a meeting that has been attended, the user is browsing his recent visit to a museum or cultural event and can also allow others to browse these experiences. A further step would be to allow real-time remote participation of a friend or relative (as is already done when for example a visitor uses a mobile phone to describe a place or an artwork to someone at home). Again, this can be done at various levels, including the visualization of the visitor in a virtual reality representation of the exposition room and where this virtual reality environment is made accessible on PC or other display facilities for those who couldn't join (see also [6]).

## 5. Conclusions and Future Research

From our experiences doing research on smart meeting rooms we abstracted some general viewpoints on ambient intelligence research and we took them to the area of smart museum environments. Our main observation here is that we see smart environment research in previously separate domains converge and this convergence will be beneficial for domains that until now only had ad hoc approaches to introducing intelligence in their environments.

## References

[1] E. Aarts, R. Collier, E. van Loenen & B. de Ruyter (Eds.). Ambient Intelligence. Proceedings First European Symposium, EUSAI 2003, LNCS, Springer, Berlin, 2003.

[2] N. Jovanovic. Recognition of meeting actions using information obtained from different modalities - a semantic approach. TR-CTIT-03-48, University of Twente, October 2003.

[3] P. Marti *et al*. Adapting the museum: a non-intrusive user modeling approach. 7th Intern. *User Modeling Conference, UM99*, J. Kay (ed.), Springer, New York, 1999.

[4] P. Marti *et al*. Situated Interaction in Art Settings. Paper presented at Workshop on 'Situated Interaction in Ubiquitous Computing' at *CHI 2000*, April 3, 2000.

[5] I. McCowan *et al*. Modeling human interaction in meetings. Proc. *IEEE ICASSP 2003*, Hong Kong, 2003.

[6] A. Nijholt. Gulliver Project: Performers and Visitors. *EVA 2002* Electronic Imaging & the Visual Arts. V. Cappellini, J. Hemsley, G. Stanke (eds.), Pitagora Editrice Bologna, 241-245.

[7] A. Nijholt. Multimodality and Ambient Intelligence. In: *Algorithms in Ambient Intelligence*. W.F.J. Verhaegh, E.H.L. Aarts & J. Korst (eds.), Kluwer, Boston, 2003.

[8] M. Trivedi et al. Active Camera Networks and Semantic Event Databases for Intelligent Environments. Human Modeling, Analysis and Synthesis, Hilton Head, SC, June 2000.

[9] J. Umiker-Sebeok. Behavior in a museum: A semio-cognitive approach to museum consumption experiences. *Signifying Behavior*, Vol. 1, No. 1, 1994.

[10] E. Veron & M. Levasseur. *Ethnographie de l'exposition*. Paris: Bibliotheque Publique d'Information, Centre Georges Pompidou, 1983.