ANTON NIJHOLT
University of Twente
Centre of Telematics and Information Technology
Human Media Interaction Research Group
P.O. Box 217, 7500 AE Enschede, The Netherlands
anijholt@cs.utwente.nl

Human and virtual agents interacting in the virtuality continuum

1 Introduction

In this paper we take a multi-party interaction point of view on our research on multimodal interaction between agents in various virtual environments: an educational environment, a meeting environment, and a storytelling environment. These environments are quite different. All these environments require the modeling of multimodal interaction: interactions between human users, the environments and objects represented in the environments, and embodied conversational agents that represent human users or that have been designed to play particular roles in the environment. E.g., such agents play the role of an information or navigation agent, play the role of a meeting assistant in a virtual meeting environment, or play the role of an actor in a virtual storytelling environment. Rather than interacting with one particular user they need to interact with different human and synthetic agents, they need to know about the properties (personalities, intelligence, emotions, capabilities, etc.) of these different agents, they need to know who is aware of what they are saying or doing, and they need to maintain a model of the multi-party dialogue between the different agents. Presently, no one is able to provide such a model of multi-party interaction. On the other hand, more and more we see research projects touching on these topics and for that reason we think it is useful to make our attempts to enter this area of research explicit.

In this paper we introduce several projects we work on. The main aim to introduce these projects in one paper is that it makes it possible to compare research approaches on interaction modeling in virtual, mixed-reality and real (physical) environments. For these environments, interaction modeling means multimodal (verbal and nonverbal) interaction modeling and it means (1) modeling human behavior and human – human interaction behavior, (2) modeling human and virtual actor behavior and human – virtual actor behavior, and (3) modeling behavior of virtual actors and virtual actor – virtual actor behavior. In addition, see also [15], it is possible to talk about different types of audience involvement, that is, virtual and real people that mainly observe but may be given the opportunity to influence the activities by their presence and their reactions.

When talking about virtual actors we can as well talk about smart objects, smart graphics, and interactions with a smart or ambient intelligence environment. In fact, we are looking at the so-called virtuality continuum as mentioned in [13], but rather than confine ourselves to this continuum in terms of display technologies, we take a much broader point of view, in particular the point of view of interaction made possible by, among others, these display technologies. Speech and language are among the modalities that we have considered, but also haptics, facial displays and animations.

Rather than introduce new research results we choose to give an overview of the multi-party interaction projects we work on, with the aim, both for ourselves and the readers, to make our intuitively available framework more explicit and to give it a more fundamental basis by comparing it with research projects performed by others.

A few words about the remainder of this paper. In this introduction we have set the stage. This paper is about the integration of multimodal interactions, smart and virtual environments, ambient intelligence and human and virtual actors. In order to do so, apart from this introduction, we discuss several projects we started or got involved with in recent years. To discuss these projects in one framework we need to explain several of the concepts mentioned above. These are virtual actors (also called virtual humans or embodied conversational agents), multi-modal interaction, smart environments (or ambient intelligence) and smart objects, virtual reality environments, virtual storytelling, and the virtuality continuum. In addition we have to explain about modeling multi-party interaction in the virtuality continuum. This explanation is done in the different subsections of section 2. In section 3 we discuss how these concepts play a role in our research projects. Again, the issue of verbal and nonverbal interaction between human and synthetic agents in the continuum between real and virtual environments is our starting point of observations. In section 4 we will more explicitly zoom in on the issue of multi-party interaction in the virtuality continuum. Section 5 provides a view on our projects from the virtuality continuum perspective. Section 6 contains a short discussion and observations about future research.

2 Actors and Environments

This section is about the different concepts that are relevant in modeling verbal and nonverbal multi-party interaction in the virtuality continuum. We don't want (or are able) to present a formal definition of this continuum, but intuitively it will be clear that we can go from fully real (physical) environments with no computer-generated stimuli to immersive virtual environments where all stimuli are computer generated. In between there are concepts denoted by (virtually) augmented reality or (realistically) augmented virtuality. Examples: (1) a human can interact with a virtual environment and with one or more virtual humans that inhabit this environment, (2) a human can interact with virtual and physical environment and virtual and real humans when devices and modalities allow a smooth integration of these realities, and (3) a human can interact with one or more other humans in a particular environment. Below we discuss the different concepts that play a role.

2.1 Virtual Actors

Virtual actors or virtual humans are two- or three-dimensional human-like animated characters that show intelligence, emotions and that know how to interact with human users. They can be used for tutoring or training tasks or to provide information and demonstrations. In recent years virtual humans have been designed for various tasks, among others to provide information services, to help users to navigate in virtual and web environments and to play the role of a virtual tutor, helping a student to perform a certain task. Many examples of research in this area can be found in [14]. Since these virtual actors appear on the screen or in the environment in an embodied way (that is, as an animated 2D or 3D human-like face or figure) they not only allow verbal (speech and language), but also nonverbal communication behavior, displayed through gestures, body and head movements, gestures and facial expressions, visible for the user of the system, or rather the human partner of the embodied agent.

2.2 Multimodal Interaction

Multimodal interaction is about the integration of modalities. When a user interface allows different input modalities these modalities need to be integrated. Keyboard, mouse, touch screen, audio (speech), video (vision) and haptic (force feed-back) recognition are among the possible input modalities of a computer system. Location sensors, cameras and microphones make it possible to know about the position of the user in the environment, the user's body posture and body movements, what he or she is doing with his or her hands, the head movements and the facial expressions and they make it possible to track the user in the environment. In traditional human-computer interaction there is one user and one computer (monitor, keyboard, mouse, camera, haptic device, etc.) allowing face-to-face interaction. When a user employs different modalities during the interaction, the system needs to integrate them. The fusion of information coming from different input modalities makes it possible for a computer system to understand what a user wants by disambiguating 'utterances' in one or more particular modalities using information coming from other modalities. Multimodal interaction is also about the fission of information. What combination of modalities does the computer system choose to react on input from the user or to draw the user's attention? When embodied conversational agents are used in the interface human-like verbal and nonverbal communication modalities become possible.

2.3 Smart Environments

The third concept we mentioned is smart environments and smart objects. More than traditional human-computer interfaces smart environments allow the detection and anticipation of a user's actions. In fact, rather than talking about a user's actions, it is more appropriate to talk about events taking place in an environment and about activities of users or inhabitants in a particular environment. Smart environments require sensors to perceive what its inhabitants are doing. Clearly, we know about microphones and camera's, but these environments can also be equipped with sensors that sense movement, location and other, less global, activity; e.g., using a keyboard or touch screen. Tracking technology makes it possible to follow and anticipate actions in a smart environment. Ubiquitous and Pervasive Computing are among the terms that are used to identify the research objectives to equip environments with communicating sensors and embedded computing devices that allow distributed intelligence. Ambient Intelligence is the term that is used to indicate the combination of Ubiquitous Computing and Intelligent and Social Interfaces. Ambient Intelligence follows the remark of Mark Weisser [24]: "The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it." and, added by Norbert Streitz: " . . . and facilitate a coherent and social experience when interacting and cooperating within the environment by providing appropriate affordances." Intelligence, whether rational, emotional or social, needs to be embedded in the environment, requiring embedding in walls, objects, clothes, and humans.

2.4 Virtual Reality Environments

We assume that readers know about virtual reality. When we talk about virtual reality we mean 3D graphical environments that allow a visitor to take every viewpoint in the environment, when equipped with the appropriate devices, has an immersive experience in this environment. However, it is also possible to talk about desktop virtual reality environments or virtual environments that allow a 3D or almost 3D experience while using simpler and more accessible display facilities. For example, it is not always useful to allow the user access to a particular environment or to allow the user to change the environment. Some applications just require that the user experiences the environment, in other cases it may be sufficient that the user gets the opportunity to interact with the environment, or a particular embodied agent present in he environment, without being able to access the environment or to make any changes (except for asking, inviting or provoking certain verbal and nonverbal utterances of the agent) to the environment. Obviously, in a virtual environment we can have several virtual inhabitants, that is, virtual humans that display activities and interaction behavior. One of more of these inhabitants can represent human actors that in real-time participate in the activities or that have sent their virtual representations (avatars) to take care of their interests.

2.5 Virtual Storytelling

Virtual storytelling is the next concept we want to introduce here. In virtual storytelling environments we don't have actors that play a role, rather we have actors that play themselves, brought together in a particular context and pursuing certain goals. A goal may be to survive in a world co-inhabited by predators, to kill the villain who has captured the princess, or to make optimal profit in a world with competing companies. Automatic storytelling can be done according scripts, scenarios and story grammars that restrict the behavior of the virtual actors or there can be a supervisor (director) that guards the development of the story, taking care of structure, consistency and maybe even, when required by the application, unexpectedness, suspense and believability of the actors. Plot construction is obtained from the actions of actors that pursue their goals in an environment that puts constraints on these actions. The virtual actors are autonomous agents obeying contextual constraints. There is, of course, interaction

between the virtual actors in a virtual storytelling environment. However, the interaction, at least in the current generation of virtual storytelling environments, is more important at the level of (the description of) virtual physical behavior and activities than at the level of verbal and nonverbal utterances exchanges. The plot that develops can be given to a narrator agent that determines how to structure the plot for presentation and finally a virtual presenter can be made responsible for telling the story. Rather than telling the story we can also have virtual drama where the agents become embodied and act in a play that is visualized in a 3D virtual reality environment. In that case the narrator becomes a play writer and there is no need for a presenter. Virtual storytelling can be made interactive by allowing the human user to act as a director during plot development or to allow the human user to control one of the virtual characters, that is, the virtual (embodied) character represents a human player in the play.

2.6 Virtuality Continuum

Finally, we want to mention the concept of the virtuality continuum as it was introduced in the literature [13]. In fact, in previous subsections we already alluded to this concept. In Figure 1 we illustrate this continuum from full reality to full virtuality.

Figure 1: The virtuality continuum

From left to right there is an increasing degree of computer-produced stimuli. At the extreme right we have immersive virtual environments where all stimuli are computer generated. We can also look at this continuum from the point of view of smart environments and multi-party interaction. A real environment can have human inhabitants that interact with each other. They see and hear each other and can understand the interactions taken place and the behavior of the inhabitants of the environment. Turning this environment into a smart environment requires the distribution of (smart) perceptory devices throughout the environment allowing the environment to keep track of locations, activities and interactions and provide support, anticipating what the inhabitants need. Making use of these devices smart objects and smart interface agents can be designed that can both address human inhabitants of the environment and that can be addressed by the human inhabitants of the environment. Agents can get human-like embodiment, that is, we can have virtual humans taking part, using appropriate display technology, in activities and collaboration with human users and inhabitants of the environments. These virtual humans can either be fully synthesized and autonomous, they can real-time represent humans that remotely visit, employ or work in the environment, or they can be something in between. And, of course, they can be made to interact with human partners in the environments. This allows us to look at the mixed reality continuum from a (multi-party) interaction point of view. We will return to the virtuality continuum in section 4.

3 Multi-modal Interactions between Human and Virtual Actors

In this section we introduce several of our projects on multimodal interaction in the virtuality continuum. The first project is also historically our first involvement with virtual reality environments and interactions with those environments, including interactions with virtual actors (avatars) available in these environments. This is the so-called Virtual Music Centre (VMC) project. We start with this project because the environment has been set up, maybe too ambitiously at that time, to allow any kind of (virtual) human – (virtual) human interaction, allowing both verbal and nonverbal interaction, multi-user use of the environment, distributed access to the environment, and personalized access to the information available in the environment. The second project we introduce is the INES (Intelligent Nursing Education System) project. In the interface of INES we model the interaction between three agents: a student performing an exercise, a virtual patient, and a virtual tutor. The third project we mention is the AMI (Augmented Multi-party Interaction) project. Here we need to model the (verbal and nonverbal) behavior of a group of meeting participants in a smart environment in order to supply, among others, real-time support to the meeting participants. Finally, we mention our work on the Virtual Storyteller. In this environment, rather than having agents exchange information, we have synthetic agents put together in the same environment and while they act in accordance to their goals and emotions a story (or play) develops.

3.1 VMC: The Virtual Music Centre Project

Some years ago we built a virtual theatre environment in VRML [14]. The theatre was built according to design drawings of the architects of the building. Visitors can explore this desktop environment, go from one location to another, ask questions to available agents, click on objects, etc. Karin, the receptionist of the theatre, has a 3-D face that allows simple facial expressions and lip movements that synchronize with a text-to-speech system that mouths the system's utterances to the user. Karin was in fact the front-end of a dialogue system that allowed users to interact with the system about theatre performances, available tickets and make reservations. Other agents in this environment have been introduced. One example is a navigation agent, which knows about the building and can be addressed using speech and keyboard input of natural language. The visitor can ask about existing locations in the theatre and when recognized a route is computed and the visitor's viewpoint is guided along this route to the destination. A Java based agent framework was introduced to provide the protocol for communication between agents. It allows the introduction of other human-like agents. Some of them are represented as communicative humanoids, more or less naturally visualized avatars standing or moving around in the virtual world and allowing interaction with visitors of the environment. In a browser that allows the visualization of multiple users, other visitors

become visible as avatars. The main goal of this environment, only partly realized at that time, was to allow any visitor to communicate with agents and other visitors, whether visualized or not, in his or her (virtual) view. That means we can have conversations between agents, between visitors, and between visitors and agents. In a multi-user version of the VMC we were able to have several visitors represented by 3D avatars in the environment. The end result was an environment where different types of agents were available and different kinds of interaction between agents were designed. In [15,16] we have more detailed observations about this environment and the different ways agents can play a role. The environment and its inhabitants and visitors were not agent-based. That was the main reason that despite the wonderful visualization of the environment and the intelligence of some of the agents not to continue this project and turn our attention to agent-based systems in virtual reality environments.

3.2 INES: Multi-party Interaction in Educational Environments

The INES (Intelligent Nursing Education Environment System) is an application we designed that allows students to use multimodal interaction – including speech and haptics - with a virtual embodied tutor and a virtual embodied patient. The environment is meant to teach procedural tasks, e.g., as has been implemented in our system as a first example, to give the virtual patient a subcutaneous injection. This task requires the execution of several subtasks, for example, taking care that the instruments are sterilized, that there is communication with the patient, and that the injection is done in a correct way [6]. The student has to master this nursing task and therefore has to communicate, using multimodal input (haptics, speech and keyboard), with the environment and its inhabitants, that is, the virtual patient and the virtual tutor. The virtual tutor monitors the student and provides affective feedback in order to smoothen the learning process. In this situation we have to model the interaction between student, patient and tutor, using speech recognition, speech synthesis and haptic and keyboard input. The tutor agent monitors the student, knows about the task that has to be performed and knows when a student makes errors. Depending on the seriousness and the frequency of the errors a student makes it decides to let the student continue (let the student learn from his or her errors), to suggest better and more useful approaches to the problem the student is trying to tackle or to provide a demonstration, showing the student what has to be done. Obviously, in this situation the student knows about the tutor and the patient, and the student knows about the tutor and the patient. The patient mainly just reacts on what the student is doing. Unlike what was made possible in the VMC, there is no multi-user version, there are only very restricted spoken dialogue systems and there is much less freedom for the user or visitor to exploit and explore the environment. However, much more than in the VMC it has become possible to model three-way verbal and nonverbal interaction between agents (including a human) that can have knowledge about each other, know what the other is doing, and know about the environment and the tasks that have to be performed. For example, in the current system, the tutor agent makes assumptions about the emotional state of the student [4] and the student can look at the facial expressions of the tutor. The student can ask the virtual patient to change his position. The patient may ask the student for explanation. Although the modeling of individual agents is not that deep, the agent-based design allows adding more capabilities to the individual agents without having to redesign the full system [5]. Current research aims at providing the tutor agent also with information about the emotional state of the student by capturing the student's facial expressions by a camera.

3.3 AMI: Augmented Multi-party Interaction in Meeting Environments

The third project we want to discuss is on meeting modeling [11,17]. How can we model what is going on during a meeting? In particular we look at the two European Union (EU) projects in the area of Information Society Technologies (IST) that we are involved with. These projects are M4 (Multi-Modal Meeting Manager), part of the 5th framework programme of the EU, and AMI (Augmented Multi-party Interaction), part of the 6th framework programme. In meetings several people are involved. The aim of these projects is to model what is going on during meetings, both to give real-time support to the meeting participants during the meeting and to allow off-line retrieval of information (e.g., to ask for a summary or to ask for a list of the important decisions that were made) from a particular meeting. Clearly, during meetings we have multi-party interaction. That is, meeting participants discuss the meeting topics and apart from their verbal contributions to the meeting they are also engaged into nonverbal communication with the other meeting participants. Just as the verbal contributions to the discussions, these nonverbal contributions need to be interpreted (as it will be done by other participants) taking into mind the roles the different participants have in the discussion, their backgrounds and their personalities. And, in order to be able to interpret what is going on during a meeting, and, in order to be able to give real-time and off-line support, these verbal and nonverbal activities of meeting participants and the meeting participants themselves need to be modeled. Although in the European context the M4 and the AMI project are two large projects, involving almost hundred researchers, it is certainly clear that to reach the ambitious aims need more time than have been made available. Clearly, the projects can be considered as an attempt to make the issues clear that are important in the semantic and pragmatic interpretation of verbal and nonverbal multi-party interaction in a particular environment. This modeling has to take into account spatial aspects, the fusion of information coming from multiple input sources and from different people that inhabit the environment.

3.4 Virtual Storytelling Environments

The next project on multi-party interaction that we want to introduce is our virtual storytelling project. It is very different from the previous projects; on the other hand, it is about agents - rather autonomous agents – that are put together in one environment where they pursue their goals. Since they are in the same environment we can have conflicts and cooperation. In current storytelling systems there is hardly modeling of verbal and nonverbal interactions between (semi-) autonomous agents that meet in a storytelling environment. Our system [21,22] is no exception. What we have is agent interaction behavior at a rather abstract level of activities. For example, the prince protects a princess, the villain abducts the princess, and the prince kills the villain. Prince and princess are happy. Being able to provide agents in a storytelling environment with beliefs and desires such that such a story develops is

already far from trivial. More human-like properties can be modeled in the agent model in order to get more detailed, consistent, well-structured and natural stories. In our system we introduced three extras: a virtual director, personality characteristics of agents, and an emotion model on top of the beliefs and goals of the agents in the environment. The emotion model allows the appraisal of events from an emotional point of view. An agent may decide to kill or to flee in a particular situation. This may depend on the agent's ability to kill (available weapons, strength), its emotions (being very angry) and its personality (turning angriness into an aggressive action or into some kind of self-reflection). The virtual director in our environment guards the development of the story. It can disapprove of an agent's intended action because it would destroy a believable story. For that reason it can forbid that particular action or it can introduce an unexpected obstacle in the environment making it more believable that an action cannot be performed.

3.5 Multi-party Interaction Modeling in the Virtual Reality Continuum

In the previous sections we considered several types of multi-modal and multi-party interactions. What kind of theories and models are available? In a physical environment where human activities are captured using sensors and multi-sensory media interpretation (an ambient intelligence home environment, a smart office environment, a changing mobile environment) we need to be able to model what is going on during a multiparty interaction in a particular context where all members of the party are human. In the virtual reality continuum it requires multimodal interaction between participants where participants are human, synthetic or some kind of mixture between human and synthetic. Whether the participants are human or virtual is not that important from a modeling point of view. We have to deal with a fusion of information coming from different media sources and with a fission of information to different presentation modalities. These modalities are human: speech, natural language, facial expressions, gestures, body postures and gaze directions. In order to understand what is displayed using these modalities we need models from which such displays can be generated. That is, models that allow translation of information display to individual modalities, that allow to choose between modalities and that distribute information display among modalities. Obviously, instantiations of these models should also allow to be filled in by information coming from different modalities in order to build up a representation of what has been going on in multiparty interaction. We may want to assume, depending on the task domain, that participants are cooperative and make attempts to have a good dialogue. This allows us to predict a next interaction act of a human participant and to allow a more complete interpretation of this act and to generate responses using the appropriate modalities of the virtual participants.

4 The Virtuality Continuum Revisited

As mentioned before, the virtuality continuum was introduced by Milgram and Kishino [13]. It has been accepted, commented upon and new technologies and ideas since the paper's appearance in 1994 need to be taken into account when we take its viewpoint while looking at current state-of-the-art projects. In the period following the publication of this paper we have seen many developments in graphics, virtual reality and interaction technology that allowed the design and building of environments and applications on different positions in the continuum. In addition, rather than having a one-dimensional continuum it seems to be more useful to consider a multi-dimensional continuum with different viewpoints and extremes along the axis. In the original paper the emphasis is on mixed reality visual displays. Other dimensions were mentioned, but rather as derivatives from display possibilities. For example, what can we display of the real world in the world knowledge dimension, or how does the visual display help to realize presence in the extent of the presence metaphor dimension.

In the previous sections we made sufficiently clear that more dimensions need to be considered. Apart from caves, virtual collaborative work and multi-user environments, virtual entertainment environments and virtual workbenches have been introduced. In research on smart environments and ambient intelligence smart objects are introduced that sometimes need display and other interaction technology. In these environments it is a rather natural step to have mixed reality. People interact with smart objects that display physical properties that are not really there, real objects having virtual intelligence and emotions can be introduced in such environments (e.g., an AIBO or an other robot) and virtual humans can be displayed that show autonomous or semi-autonomous behavior or just represent in real-time the behavior of a human visiting the environment.

To make these smart objects, robots, and virtual humans useful, social or entertaining we need to be able to interact with them and this should become possible in such a way that they live up to our expectations concerning their behavior, and their emotional, social and logical intelligence. That is, we want them to take part in activities with their human partners in these environments. Hence, among the dimensions that need to be employed nowadays are the virtual humans and the multimodal interaction possibilities in a virtuality continuum. There is more, for example we can talk about the role of sound and smell in environments and about the architecture of the environment. How familiar can we make it, how can it represent or acquire the personality of its main inhabitants, how does the environment learn and adapt and how does it allow its inhabitants to make changes?

Other issues that need to be dealt with are the physical constraints that the environment needs to have. An environment that needs to resemble reality also needs to mirror real-world physical laws. However, there may be good reasons to allow different laws. Indeed, an educational environment meant to learn physical laws needs to be able to represent them. But also in this case it might be quite informative to allow non-realistic variations. In an entertainment environment deviations can enhance the entertainment value. In an environment where we have both real and virtual humans we may improve physical laws by allowing agents to perceive more of their environment and their fellow participants than is possible in the real world. As an example, do we want to take into account hearing distance when two virtual agents in a virtual reality environment need to communicate with each other? Or, similarly, do they need to be able to see each other before taking actions to meet or ignore each other? That is, for seeing and hearing it is not always necessary to take into account real-world physical constraints; we can also decide to design enhanced behavior with respect to such properties. Holographic images can have different properties, but also

human inhabitants of smart environments or humans communicating with smart environments can have implants that take over or enhance cognitive functions.

What need to be modeled when we put human and virtual agents in the virtuality continuum? Many issues need to be discussed: appearance, behavior, intelligence, emotion display, gestures, facial expressions, posture, etc. However, this external behavior can be generated, using graphics and animations, from internal models of individual agents and from models of interaction between agents, whether they are human or synthetic. Assuming being able to display and animate virtual humans in a sufficiently believable way, we should concentrate on the role of interactions between humans, humanoids that represent and mimick humans and semi-autonomous and fully autonomous agents that inhabit worlds somewhere positioned along the dimensions of the virtuality continuum.

Without making any distinction, for the moment, between the different types of agents that can inhabit a world somewhere along the dimensions of the virtuality continuum, it is clear that we need models of multi-party interaction (cf. [23]) 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 meeting participants. 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 [19]. 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 [9]. 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) [12].

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 [7] 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 [2].

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 computer-supported collaboration during a design process, are we talking about real-time or off-line meeting support, or is the system employed in a home environment? Clearly, in an office environment people behave differently than when they are in their home environment. Understanding what is going on in a particular environment (allowing real-time support and off-line retrieval) 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 office situations, in home situations and in public spaces [1,8].

It is certainly not our intention here to survey all existing agent theories that we expect to be useful in the context of the virtuality continuum. However, from our observations it should be sufficiently clear that when we introduce human and virtual agents in the virtuality continuum, the above-mentioned aspects have to be dealt with in our models and designs of worlds in the virtuality continuum.

5 Back to the Virtual Music Centre, INES, Storytelling, AMI and towards a Virtual Meeting Room

In this paper we have put our research in the perspective of the VR continuum. Although all the projects or our anticipated involvement in the projects are assumed to contribute to this VR continuum, it has not been possible to cover this continuum in a systematic way. Maybe more projects concerned with virtual reality interactions are needed in order to get a more comprehensive view of the field. We spend a few words on differences and common properties of the projects mentioned above.

The most interesting and ambitious environment we worked on is the Virtual Music Centre. In retrospect we have to conclude that researchers were not ready to elaborate the many issued that need to be worked on in a systematic way. Clearly, with more advanced technology that would make it possible to introduce modules taking care of

particular properties of virtual humans, interactions and environments at a sufficiently believable level, the situation would have been different. Issues that were addressed were speech recognition, (multimodal) dialogue modeling, natural language processing, multi-user environments, virtual reality, embodied agents, animations and navigation. One of the issues that was not dealt with in the first phase of the project was the internal modeling of agents and the embedding of the individual agents in a multi-agent framework.

Our INES system allows three-way communication between a student, a virtual patient and a virtual tutor. Rather than starting with individual agents, here the starting point was an agent framework allowing the sending of messages between agents about their activities, where the agents act either as interacting software modules, not visible to the users, or as agents that act in the framework with the many other agents, but also need to interact with a human user of the system. The tutor and the patient are represented in virtual reality as virtual humans. The main shortcoming of this educational environment are not so much the framework of interacting agents but the limited ways of being able to set up dialogues with the agents and the knowledge each agent has about interactions the user (student) has with the other interface agent. In the INES system the virtual tutor makes assumptions about the emotional state of the student based upon the performance of the student. The teaching strategy of the tutor can be adapted to the characteristics of the student. However, there is quite an imbalance when looking at the properties and capabilities of these three agents (student, tutor, virtual patient).

It is difficult to compare our virtual storytelling environment with the other environments discussed in the previous subsections. The main reason is that in the storytelling environment the agents are on the one hand much more autonomous and much less intelligent than the agents modeled in the VMC or INES environment, while on the other hand their autonomy is decreased by the pre-authored narrative (or the virtual director) and their intelligence is increased by these narrative structures. Clearly, a story is heavily character-based, it is created by the actors, their goals and their emotions, but this alone does not guarantee an interesting story. Things need to go wrong, dilemmas need to be introduced, there should be surprise and a build-up of tension during the main part of the story, and therefore at every moment the possible interactions between the actors are limited. Hence, looking back, here we have an environment where, at least at the moment, we have no visualization and no embodied agents, but where the behavior of our agents is determined by their internal model of believes, desires, intentions and their emotions, and where the actions of the agents are also dependent on their personality model.

In these projects we probably have illustrated many of the issues that need to be tackled when attacking the dimensions of the virtuality continuum. What is missing, and is being attacked now, is our work in the AMI (Augmented Multi-party Interaction) project. As mentioned, in the AMI project the aim is to understand what is going on in a meeting room. Understanding has to be done by computers that are fed by input coming from cameras and microphones (and maybe input from electronic whiteboards, notebooks, available agenda, participant lists, etc.). Being able to understand allows real-time support and it allows intelligent browsing (including retrieval and summarization) of what has happened in the environment. This browsing requires multimedia output of results.

There is no other way to understand what is going on in a meeting room than to understand what is going on between the participants in the meeting. Rather than being able to model the interaction between a computer and a user we need to be able to model the verbal and nonverbal interactions between the participants (and maybe between the participants and the environment) and this means we need to look at all the aspects of social and emotional relationships mentioned in the previous section of this paper.

Being able to recognize, using cameras and microphones, what is going on in a particular environment allows not only off-line browsing, retrieval and summarization, but also a virtual reality representation of the activities that took place. This virtual reality representation visualizes the environment, but it also visualizes through avatars the participants of the meeting and their activities. It allows paying a visit to a meeting in the past, either as a former participant of that particular meeting or as someone who wants to get a feeling of what was going on during that particular meeting. Such an outsider can view the meeting from the point of view of an audience member, but he or she has also the opportunity to attend the meeting from the viewpoint of a particular participant. Even more interesting is the possibility to have a real-time translation of everything that is going on in a meeting to events that take place in a corresponding virtual reality environment. This can be useful for the meeting participants, but, more importantly, it also allows real-time participation of remote participants and autonomous agents. A remote participant can be represented by its avatar, being visible for the other meeting participants. The remote participant can attend the meeting by controlling its avatar and by communicating through the eyes, ears and body of its avatar. Control can be implicit since the actions of the remote participant can be real-time captured and converted into avatar actions. Depending on the importance of what is going on it may be possible to give the avatar some autonomy, acting on behalf of its owner.

We can go one step further by introducing meeting participants that are fully autonomous. So, why not have a virtual chairman without the disadvantages of a human chairman? Why not having agents that act as meeting assistants? For example, Neem is a project of the University of Colorado [3] that aims at introducing different intelligent agents in a distributed business meeting environment. These agents have to assist the meeting participants. In Neem three agents are considered: an informing agent (assisting in obtaining necessary information, e.g. through a web search), a social agent (helps to build common ground) and an organizational agent (keeping track of time, etc.). Underlying their behavior is Bales' Social interaction Systems theory [1] and organizational theories of problem solving.

6 Discussion

In this paper we have looked at the virtuality continuum from the viewpoint of embodied agents that a user can interact with and that take part in activities in smart and virtual environments. We certainly did not present a comprehensive overview of the research that is going on or all the issues that play a role. A recent paper in which we

discuss issues of the impact of the environment on its inhabitants is [18]. A preliminary discussion on real-time transformation from activities in a smart meeting room to corresponding activities in a virtual room can be found in [20]. Current research activities aim at making the translation of verbal and nonverbal communication between humans in a physical environment to virtual humans acting in virtual environments possible. Future research activities aim at allowing real-time participation and representation of remote meeting participants in a virtual meeting room. Interfaces that allow a smooth transition between real and virtual environments (see e.g. [10]) need to become subject of research.

7 References

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

R. Conte & C. Castelfranchi. Simulating multi-agent interdependencies: A two-way approach to the micro-macro link. In: *Social Science Microsimulation*, K.G. Troitzsch, U. Mueller, N. Gilbert & J.E. Doran (eds.), Springer, Berlin, 1996.

C. Ellis & P. Barthelmess. The Neem dream. Proceedings *Tapia '03*, October 2003, Atlanta, Georgia, USA, 23-29.

D. Heylen, M. Vissers, R. op den Akker & A. Nijholt. Affective feedback in a tutoring system for procedural tasks. *Affective Dialogue Systems*, LNCS 3068, E. André et al. (Eds.), Springer, 2004, 244-253.

D. Heylen, A. Nijholt and R. op den Akker. Affect in tutoring dialogues. *Journal of Applied Artificial Intelligence,* February 2005, to appear.

M. Hospers, E. Kroezen, A. Nijholt, R. op den Akker & D. Heylen. An agent-based intelligent tutoring system for nurse education. Chapter 9 in *Applications of Intelligent Agents in Health Care*, J. Nealon & A. Moreno (eds.), Whitestein Series in Software Agent Technologies, Birkhauser Publishing Ltd, Basel, Switzerland, 2003, 143-159.

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, 135-149.

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

J. Gratch & S. Marsella. Tears and fears: Modeling emotions and emotional behaviors in synthetic agents. In Proc. of the Fifth International Conference on *Autonomous Agents*, Montreal, Canada, ACM Press, 2001, 278-285.

B. Koleva, H. Schnadelbach, S. Benford and C. Greenhalgh. Traversable Interfaces between Real and Virtual Worlds. In: Proceedings of *CHI 2000*, The Hague, 233-240.

McCowan, D. Gatica-Perez, D. Bengio, D. Moore & H. Bourlard. Towards computer understanding of human interactions. *Proc. European Symp. on Ambient Intelligence (EUSAI)*, LNCS 2875, Springer, Berlin, 2003, 235-251.

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*. NY: Guilford, 51-87.

P. Milgram and F. Kishino. 1994. A Taxonomy of Mixed Reality Visual Displays*. IEICE Transactions on Information Systems.* vol. E77-D, no. 12, Dec. 1994.

Nijholt and J. Hulstijn. Multimodal Interactions with Agents in Virtual Worlds. Ch. 8 in *Future Directions for Intelligent Information Systems and Information Science*, N. Kasabov (ed.), Physica-Verlag, 2000, 148-173.

Nijholt. Towards virtual communities on the Web: Actors and audience. Proc. Intern. ICSC Congress *Intelligent Systems & Applications (ISA'2000)*, Vol. II, F. Naghdy et al. (eds.), ICSC Academic Press, Canada, 725-731.

Nijholt. From virtual environment to virtual community. In: *New Frontiers in Artificial Intelligence.* T. Terano et al. (Eds.). LNCS 2253, Springer Verlag, Tokyo, 2001, 19-26.

Nijholt, R. op den Akker and D. Heylen. Meetings and meeting modeling in smart surroundings. In: *Social Intelligence Design*. A. Nijholt & T. Nishida (eds.), 3rd Intern. WS, CTIT Series WP04-02, Enschede, 2004, 145-158.

Nijholt. Where computers disappear, virtual humans appear. *Computers and Graphics*, Vol. 28, No. 4, Elsevier, 2004, 465-476.

Ortony, G.L. Clore & A. Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, 1988.

R. Poppe, D. Heylen, A. Nijholt & M. Poel. Towards real-time body pose estimation for presenters in meeting environments. Submitted for publication.

M. Theune, S. Rensen, R. op den Akker, D. Heylen & A. Nijholt. Emotional characters for automatic plot creation. In: *Technologies for Interactive Digital Storytelling and Entertainment.* Second Intern. Conf., TIDSE 2004, Darmstadt, Germany, June 2004, S. Göbel et al. (Eds.), LNCS 3105, Springer, Berlin, 95-100.

M. Theune, S. Faas, D. Heylen & A. Nijholt. The virtual storyteller: Story creation by intelligent agents. TIDSE 03: *Technologies for Interactive Digital Storytelling and Entertainment*, S. Göbel et al. (Eds.), Fraunhofer, 2003, 204-215.

D. Traum & J. Rickel. Embodied agents for multi-party dialogue in immersive virtual worlds. In:. Proc. 1 Int'l Joint Conf. on *Autonomous Agents & Multi-Agent Systems* (Vol. 2), 2002, 766-773.

M. Weiser. The Computer for the Twenty-First Century. *Scientific American*, September 1991, 94-10.