Available online at www.sciencedirect.com
SCIENCE DIRECT
Computers & Graphics 28 (2004) 467–476
COMPUTERS & GRAPHICS
www.elsevier.com/locate/cag

Where computers disappear, virtual humans appear

Anton Nijholt*

Department of Computer Science, Twente University of Technology, P.O. Box 217, 7500 AE Enschede, The Netherlands

Abstract

In this paper, we survey the role of virtual humans (or embodied conversational agents) in smart and ambient intelligence environments. Research in this area can profit from research done earlier in virtual reality environments and research on verbal and nonverbal interaction. We discuss virtual humans as social actors and argue that, rather than is common in traditional human–computer interface research, we need to look at multi-party interaction. Virtual humans in the party need to be equipped with nonverbal communication capabilities, including the display of emotions.

Keywords: Virtual humans; Embodied conversational agents; Ambient intelligence; Smart environments; Virtual reality

1. Introduction

An important issue in traditional and current human–computer interaction is the design of the interface. One current and future issue in human–computer interaction is the disappearing computer [1]. Where is the interface when smart surroundings and ambient intelligence environments have become part of the physical settings? When the environment anticipates and supports our activities, how do we communicate with it? What will be visualized and which modalities for communication will be supplied? When we have a question or need support, who or what do we address in these environments? Will we not feel lost in ambient intelligence?

When every imaginable environment and its smart objects allow perception and interpretation of what its inhabitants are doing and also allows interaction between inhabitants and environment, it is certainly useful to investigate how we can design social interfaces, interfaces that emphasize human-to-human communication properties, rather than concentrating purely on designing for intelligence and efficiency. After all, we will spend much of our time in these future environments and we will spend much of that time—maybe not always explicitly or consciously—interacting with these environments.

1.1. Ambient intelligence environments

Ambient intelligence environments are intimately integrated with our everyday environments. Ambient intelligence is said to consist of ubiquitous computing + social and intelligent user interfaces allowing social interaction. This also assumes that in ambient intelligence 'the real world is the interface', the interface becomes one with physical setting. At least, that is the message that is proclaimed. Garden, house, car, sitting room, study, office and in fact every environment and its natural objects allows perception of what is going on in the environment and allows interaction by its occupants and visitors to extract and exchange information (including mood and emotions). We should feel comfortable within it, although we know that the environment has eyes and ears that observe what we are doing. We should also feel free and comfortable in addressing these environments when we need support in our activities.

As mentioned, the environments will know about us. They know about our weak and strong points, they can

*Fax: +31-534893503.
E-mail address: anijholt@cs.utwente.nl (A. Nijholt).

0097-8493/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cag.2004.04.002

be designed to fulfill affiliation needs and to induce self-disclosure since they can perform better when they know about our intimate characteristics. We even have to assume that there is a 'human audience' in the 'background'. There can be real-time involvement by those who own or provide the environment or have otherwise been hired to provide user-support. Off-line processing (manual browsing of what has been going on or automatic detection and presentation of what is in the interest of those who control the environment) is another example of audience involvement. In a home environment, we may assume that family members and friends can be given access to such browsing facilities.

Most of the current research into ambient intelligence deals with how the environment is able to identify and model users' activities, rather than how the user is willing, able or likes to communicate with the environment or have the environment communicate with him [2]. In more traditional environments multi-modality in interactions has received attention, but it has hardly been investigated how these results can be transferred to environments where the user does not always explicitly address a particular (part of a) screen or an object. Moreover, most of the research in ambient intelligence does not take into account that people may not always know who to 'talk' to and may feel the need to build some kind of relationship with an environment that supports them, observes them and keeps track of their private and job activities.

1.2. Social interfaces

A useful question to investigate is what kind of relationships do we have with our traditional interfaces? There is already a trend towards designing social interfaces, emphasizing human-to-human communication properties, rather than concentrating purely on designing intelligence and efficiency. In this research, the computer is perceived as a social actor. Interaction should be socially formed and interaction design should take into account the needs of emotions, personality, affiliation, friendship or even more. What are the characteristics of environments that can be made to help to employ the results of research into traditional interfaces? And which characteristics require new and original research?

Much of the research on social interfaces is related to the design of embodied conversational agents (ECAs). An obvious question is, will ECAs that are able to develop affinitive relationships with their human partners play an important role in ambient intelligence environments? In this paper, we have observations about natural interaction made possible by employing embodied agents in the (traditional) interface and we argue that for natural interaction in ambient intelligence environments it is useful to reserve a role for these agents in order to be able to maintain desirable characteristics of human–human communication, e.g., to establish short- or long-term relationships in and with smart environments. Embodied agents can offer intelligence, personality and emotion and therefore communication properties that help to make us feel being appreciated and that make us feel understood. This makes it possible for us to act in smart, but also in social environments. We do not argue that embodied agents resembling human beings are the only solution to the problems that may arise in smart environments. It will become clear in the next sections that embodied agents are not always necessary to allow a user to build up desirable relationships with objects or environments. Our viewpoints seem to be in contrast with approaches where users are assumed to control a smart environment by using universal service access aids implemented on remote controls and pda's. We prefer to introduce ECAs familiar with services in smart environments and acting as inhabitants of the environment than having human inhabitants and visitors struggle with remote controls. This does not mean that such devices cannot play a useful role (see e.g. [3]). But they themselves can achieve the role of a virtual personality or they can be used to display a virtual, human-like, personality on a screen to advise the user, to answer questions and to negotiate with the environment (personal assistants, mobile companions, virtual butlers). Since embodied agents (virtual humans) are particularly employed to deal with social needs and have been shown to allow for the development of interpersonal relationships, we think they may be a good starting point for research on how to deal with social communication needs in ambient environments.

1.3. About this paper

Summarizing, our aim in this paper is to investigate how useful it would be to furnish ambient intelligence environments with embodied agent technology that will help to make the environment more comfortable to its daily inhabitants and visitors, making it a better place to work, relax and have fun. In Section 2 we discuss the Computers Are Social Actors (CASA) view introduced by Reeves and Nass. In Section 3 we will discuss ECAs, survey the research issues in play and give examples of employment, in particular looking at nonverbal behavior supporting research and technology. Section 4 is devoted to the possible development of social relationships between agents and humans in ambient intelligence environments. Recent approaches from our own work are included. Section 5 contains the conclusions of this paper.

2. Actors, environments and relationships

In human–computer interaction, we can perceive social characteristics of human–human interaction. There is human-like behavior when interacting with the computer and sometimes human-like behavior is expected of the computer. Can we expect similar behavior when the user is interacting with an environment—consisting of walls, furniture, cloths and other natural objects—rather than with a desktop screen? Moreover, there is communication between these embedded computational devices allowing a much more comprehensive perception and interpretation of events taking place than is possible with a single computing device. And, also allowing much better and distributed display facilities, using audio and visualization (3D graphics, smart graphics, virtual and augmented reality). The question is how this technology contributes to the well-being of the inhabitant or how it reduces this well-being. To answer this question it helps to have a look at some observations on social reactions to information and communication technology in general.

2.1. Social reactions to communication technology

In the "Media Equation", Reeves and Nass [4] report about their experiments on human–computer interaction where humans assign human characteristics to computers. Many experiments have been done since this book was published. The results became known as the "social reactions to communication technology" (SRCT) perspective in which "computers are social actors". An example of an experiment is the following. A student is asked to sit behind a computer and to perform a particular task. When finished, the student needs to answer questions: how helpful was the computer, was it friendly, was it polite, etc. Two computers were available for answering these questions: the computer that was used for performing the task and another computer just for presenting the questionnaire and having the student answer it. It turned out that when the questionnaire had to be answered on the computer that had been used to communicate the task to the student and to help the student when performing this task, students answered much more positively and politely than when answering similar questions posed by the second computer. Clearly, people do not like to offend a computer that just tried to be helpful to them.

Many similar experiments have been performed. Computer users turned out to be sensitive for flattery and humor; moreover, they were greatly influenced, when assigning personality characteristics to a computer, by the properties of the synthesized voice in text-to-speech synthesis. And, as became clear from the experiments, it is not just a matter of contributing personality characteristics to a computer; it is also a matter of being influenced by these properties while communicating with the computer. Hence, the book's conclusion was as follows:

> 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.

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 relying on these authors' observations, we have to look at the interaction characteristics of human–environment interaction and design our own research. Just to help our intuition about the issues in play, look at remarks made by Michael Coen from MIT Labs about the effects of smart environments on their inhabitants: "The notion of being alone may disappear, or it may be changed drastically." and, "You may be in a room that's always alive and aware. And from my experiences here…when the space is 'off,' you feel it. You notice that it's not reacting. There's a void."

2.2. Environments as interaction contexts

How will humans interact with such environments? Are they able to build some kind of relationship with these environments just as they are able to build relationships with a computer that is perceived as a social actor? Or do we need to introduce explicitly visible social actors, that is, ECAs, in these environments with which users can communicate and exchange information in intelligent and social ways in order to fulfill a need to establish relationships with their environments?

Some notes are in order. The first note concerns the future. It is already the case that a large part of the professional population in Western countries spends the day with discussion, meetings and knowledge exchange and spends lots of time interacting with computers. The need to do this in the office will decrease and home, work and mobile situations will come to resemble each other more and more. Interaction forms require mixtures of efficiency, social relationship and entertaining aspects. Our hypothesis is that people prefer to be able to interact with their 'own', personalized (but not only in the current technical sense, i.e., aimed at efficiency) and therefore non-anonymous environment.

Secondly, and related to the previous observation, 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 other 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. See e.g. [5], where Gosling et al. discuss links between individuals and the physical environments they occupy and between environments and observer's impressions of the occupants of physical rooms. Similarly, we may assume that whenever technology allows, consciously and unconsciously, links are created between individuals and their (ambient intelligence) environments.

Thirdly, it is useful to distinguish between situations. Different circumstances require different kinds of interactions. Sometimes we want to see things arranged in an efficient way. Sometimes we are more concerned with a partner's satisfaction when arranging things. Sometimes arranging itself is entertaining. Both interaction and information exchange can be goals in themselves, e.g., when we enter into conversations with our children or colleagues. Efficiency is not necessarily the starting point when engaging in these conversations.

2.3. Interaction goals and public behavior

Although, the SRCT perspective makes us aware that people react socially to computers, a more detailed view can make clear many nuances. To start with, there is no such thing as *the* computer. Its performance, as it shows in the interface, can be task oriented, it can be communication oriented and it can be oriented towards establishing and maintaining relationships. In Interpersonal Theory these types are the three tracks of conversational goals [6]. The task goal in human-to-human conversation is why the conversation is started, i.e., to accomplish a certain task and part of the interaction behavior is meant to reach this goal. The communication goals aim at making the interaction process run, e.g., by allowing smooth turn taking. The relationship goals of the conversational partners set the tone of the conversation. Two broad categories of relationship goals are distinguished: communion (behaviors oriented towards connecting with one another or disconnecting from another) and agency (behaviors oriented towards exerting influence or yielding to influence). Shechtman conducted experiments to study relationship behavior during keyboard human–computer interaction and (apparently) keyboard mediated human–human interaction. In the latter case, participants used many more communion and agency relationship statements, used more words and spent more time in conversation.

So, there are differences in interaction behavior when rather than computers, humans are—or are assumed to be—involved in the interaction, and this can be a further argument to look at interaction roles and goals for virtual humans in smart environments. Since smart environments may have several inhabitants (real and virtual) at the same time, it is also useful to look at human behavior in public spaces.

Generally, in public spaces where people are aware of others, their behavior is different from that in private spaces and when they are alone. There is behavior towards others in the environment, but this behavior is not always goal-oriented or focused towards one particular person. This means also that when discussing interactions and behavior in smart environments, we should pay attention to behavior that differs from interaction in a traditional face-to-interface setting. That is, in addition to the conversational goals mentioned above, in particular the relationship goals, there are goals that apply to group behavior, behavior in a group and group culture. Hence, when being one of more inhabitants of a smart environment, where the others may be embodied agents, we also need to look at the impact on the interaction behavior and interaction preferences of its inhabitants.

When alone, people behave differently than when other people are present. When other people are present, it is useful to distinguish between focused and unfocused interaction [7]. As mentioned by Goffman [7], just being there conveys information about a person: "social attributes, his conception of himself, of the others present, and of the setting." Body idiom makes information available to others and although this can also be done in a focused way, it is usually unfocused, not aiming at an exchange of information with a particular person in the environment. Sometimes general rules of etiquette regulate such behavior. The obvious question now is whether we can observe such behavior in smart environments even when there are no other people present. Is there a difference in behavior when the room is 'off' and when the room is 'on'? Does the room function as a public place?

Knowing that you are being monitored leads to different behavior. It may make a difference when there is perceived real-time observation by a human being or that there is off-line—or the possibility of off-line—retrieval of what has been recorded. In our research, aimed at introducing embodied agents into smart environments, it is especially interesting to look at the work of Slater [8], who studied the illusion of sentience in virtual environments, with the objective to present evidence that people react to virtual characters as if they

Fig. 1. Virtual audience showing appreciation.

were real. The effect of embodied agents (or virtual humans) in face-to-face interaction has been studied before (see e.g. [9] for a short overview of research on virtual embodied tutors), but not really in a 'Cave'-like environment with experiments that involved groups of virtual humans. In these experiments, the subject was asked to prepare a 5-min talk and then, after entering a virtual presentation room and wearing a head-mounted device, was confronted with a virtual audience. Three types of audiences were distinguished: neutral, friendly and appreciative, and hostile and bored. The audience responses were triggered by a remote operator using scripts and collections of animations, pose shifts and audio effects. Examples of negative audience behavior included among others: frown, walk out of room, fall asleep and talk amongst themselves. Examples of positive audience behavior included applause, lean forwards, nod encouragingly and maintain eye contact. See Fig. 1 for a view of the positive audience.

From questionnaires and from heart rate data that was obtained it became clear that people reacted to the virtual audience in a realistic manner. It felt like giving a real presentation: heightened anxiety before talk, emotional reactions to audience, attempts to connect with, win over the audience, etc. One comment reported by Slater (given by a subject presenting to the positive audience): ''I felt great. Finally nobody was interrupting me. Being a woman, people keep interrupting you in talks much more… But here I felt people were there to listen to me.''

We conclude that when we design smart environments to be inhabited by real and virtual people we not only need to address the focused interaction between real and virtual people, satisfying several conversational goals, but also awareness and unfocused interaction behavior of virtual and real people.

2.4. Interaction visualization in ambient intelligence

This paper is concerned with the role of embodied agents (virtual humans) in smart and ambient intelligence environments. That is, when we talk about interaction visualization we talk about nonverbal communication by virtual humans. We will not discuss the issue as to whether nonverbal communication is supported by verbal communication, the other way around, or whether there is a position in between. What is clear from psychological literature that in communication content does not always play the major role. Facial expression, body language and tone of voice are often more important.

When discussing the visualization of virtual humans and their communication behavior we need to make references to available technology and how it supports natural interaction behavior. There are certainly differences between an immersive virtual environment where the user is required to wear a head-tracked, head-mounted device, a desktop environment equipped with an eye-tracker to follow the user's gaze, or a desktop computer showing a 3D VRML world or 2D animated pictures. Although no universally accepted definition for smart environments or ambient intelligence is available, it is clear that we are discussing physical environments in which computing devices are embedded in such a way that they have become invisible. The next step is the mental disappearance of computing devices. What is shown on ambient displays embedded in the environment (walls, tables, door, mobile and static objects) or rendered on physical objects should allow natural interaction with the (other) inhabitants of the environment. The environment can have its own virtual personalities, knowing about the environment and the functionalities that are available. The environment can also be inhabited or visited by a virtual human or a virtual personal assistant (virtual butler) of a human visitor. Virtual humans can be triggered to appear, using sensors that detect people in the smart environment that are in their neighborhood.

In the remainder of this paper we assume that virtual humans (in an immersive 3D representation, a 3D VRML representation or a 2D character in a desktop environment) are available and are able to communicate with users, visitors and inhabitants of our environments using verbal and nonverbal communication abilities.

Clearly, when we talk about visualization, not all modalities that can be employed in human–human or human–virtual human interaction in smart environments can be made part of the visualization. Obviously, spoken language can be shown as text or as text appearing in text balloons, there are ways to visualize intonation in spoken language, etc., but then we very much take a road that leads us away from natural interaction and the technology that is emerging from current embodied agent research.

In human–human interaction nonverbal cues play an important part in the relationship track of communication. We can ask whether we can recognize and interpret these communication aspects in human–computer and human–environment interaction and whether they can play a similar role. From the SRCT perspective we know that humans react socially to social computer behavior and having the computer display more cues about its social behavior may strengthen the social reaction. To take one extreme, there will not necessarily be a need to consider your own computer, let alone, every computer, as a personal friend with whom you want to share your feelings, using verbal and nonverbal cues. Nevertheless, there will be many situations where a user or inhabitant of an environment, will prefer communicating with an environment that allows the display of knowledge of the user, that allows display of reactive and pro-active behavior that shows understanding of the particular context of the user, including its mood and emotions. To do this we need other modalities in interaction and presenting information than just menu-based graphical user interfaces. The display of nonverbal acts in human–human interaction and human–environment interaction requires interaction models that need to underlie ambient intelligence environments inhabited by real and virtual people.

Such models are not or are hardly available. In the next sections we will discuss issues that need to be investigated in order to design comprehensive verbal and nonverbal interaction models. Nevertheless, the current state-of-the-art in designing and building ECAs, their interaction behavior and their behavior in groups and environments, allow an optimistic view on integrating embodied agents in smart environments where they know how to behave.

3. Virtual humans and nonverbal behavior

3.1. Introduction

Research on ECAs, sometimes called virtual humans, has become a well-established research area in human–computer interaction. Embodied agents are agents that are visible in the interface as animated cartoon characters or animated objects resembling human beings. Sometimes they just consist of an animated talking face, displaying facial expressions and, when using speech synthesis, having lip synchronization. Sometimes they have a graphical 3D representation, and when embodied in a virtual reality environment, they can move around, know about their environment and allow conversational partners to perceive them from different angles and to interact with them using different modalities. These agents are employed to inform their conversational partners, to explain or to demonstrate products or sequences of activities in educational, e-commerce or entertainment settings. Experiments have shown that ECAs can increase the motivation of a student or a user interacting with the system. Lester et al. [10] showed that a display of involvement by an ECA motivates a student in doing (and continuing) his or her learning task.

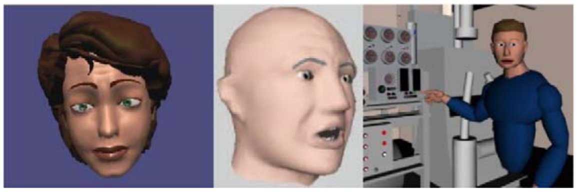

Fig. 2. Three embodied conversational agents.

Many examples of ECAs are available. In Fig. 2, from left to right, we see: Greta, designed by Catherine Pelachaud, Obi, a virtual tutor designed in our research group, and Steve, educating a student about the maintenance of a complex engine. In this example Obi shows surprise about a student's action (see also Section 4).

As mentioned, embodiment is necessary to show nonverbal communication abilities. These abilities allow more multi-modality, therefore making interaction more natural and robust. Several authors have investigated nonverbal behavior among humans and the role and use of nonverbal behavior to support human–computer interaction with the help of embodied agents in detail. See e.g. [11] for a collection of chapters on properties and impact of ECAs (with an emphasis on coherent facial expressions, gestures, intonation, posture and gaze in communication) and for the role of embodiment (and small talk) on fostering self-disclosure and trust building.

Current ECA research deals with improving intelligent behavior of these agents, but also with improving their verbal and nonverbal interaction capabilities. Improving intelligent behavior requires using techniques from artificial intelligence, in particular natural language processing. Domain knowledge and reasoning capabilities have to be modeled. Agent models have been developed that allow separation between the beliefs, desires and intentions of an agent. Together with dialogue modeling techniques rudimentary natural language interaction with such agents is becoming possible.

In the next three subsections we shortly discuss the role of facial expressions, gaze and gestures in human–human interaction and research to equip embodied agents with these capabilities.

3.2. Facial expressions

To describe emotions and their visible facial actions, facial (movement) coding systems have been introduced. In these systems facial units have been selected to make up configurations of muscle groups associated with particular emotions. Such a system should be detailed enough to describe what is happening in different regions of the face, to describe intensities and to describe the blending of emotions. Moreover, they should be detailed enough to be able to distinguish between deceptive and honest expressions. Another issue that requires encoding is the timing of facial actions. For these reasons Ekman and Friesen developed their Facial Action Coding System for scoring visually distinctive, observable facial movements.

The face has been mentioned as a primary source for obtaining information of the affective state of an interactant. However, from many experiments it has been shown that it depends on many factors (task, message, perceiver and previous experience) how weighting of different modalities is done. Modalities in the face also include movements of lips, eyebrows, color changes in the face, eye movement and blinking rate. Cues combine into expressions of anger, into smiles, grimaces or frowns, into yawns, jaw-droop, etc. For example, apart from muscle contractions in the face, fear also decreases blinking rate and head movement. Anger can show in increasing eye movement and decreasing head movement. Happiness may show in increasing blinking rate. Obviously, when using a talking face, a designer can deliberately put emphasis on particular facial actions during interaction and in this way also give more cues to the observer than is usual in real life.

3.3. Gaze

Getting a system that has natural gaze behavior involves tight co-ordination of the facial animation driver with many parameters of the dialogue manager, with the mental state of the character and its model of the user and subtle aspects of the linguistic utterance that is produced or attended to. Consider in this respect the functioning of gaze in human–human conversations [12,13]. Among others it is used to express emotion, to gain feedback from a listener, provides information about the interpersonal relationship and more generally, helps in regulating the conversation flow. For example, gazing away from or towards the interlocutor can function as an important emotional signal as well as a signal to hand over the turn or avoid the turn to be taken over. As a function in the organization of turn-taking behavior, the timing of mutual gaze (eye-contact) correlates with the information structure of the utterances (its topic-focus articulation).

In an experiment [14], we investigated the effects of different styles of gaze of Karin, one of our conversational agents, on the conversation. We had 48 subjects each make two reservations with different style versions. We videotaped the conversations, clocked the time they spent on the task, and had them fill in a questionnaire after they had made the reservations. It appeared that participants that had conversed with a version in which common gaze behavior was implemented (looking away and towards users and beginnings and ends of turns, respectively) appreciated their conversation significantly better than the other participants in most respects. They not only were more satisfied overall, they found it easier to use than a version with the minimal amount of eye movements, appreciated the personality of the agent better and thought the head movements were more natural. They were also the fastest, on average, to complete the task.

3.4. Gestures

What role do gestures play in communication and why should we include them in an embodied agent's interaction capability? Categories of gestures have been distinguished. Well known is a distinction in consciously produced gestures (emblematic and propositional gestures) and the spontaneous, unplanned gestures (iconic, metaphoric, deictic and beat gestures). Gestures convey meanings and are primarily found in association with spoken language. Different views exist on the role of gestures in communication. Are they for the benefit of the gesturer or for the listener? Gestures convey extra information [15] about the internal mental processes of the speaker: "…an alternative manifestation of the process by which ideas are encoded into patterns of behavior which can be apprehended by others as reportive of ideas." Observations show that natural gestures are related to the information structure (e.g., the topic-focus distinction) and (therefore) the prosody of the spoken utterance. In addition they are related to the discourse structure and therefore also to the regulation of interaction (the turn-taking process) in a dialogue. Apart from these viewpoints on embodiment, we can also emphasize the possibility of an embodied agent to walk around, to point at objects in a visualized domain, to manipulate objects or to change a visualized (virtual) environment. In these cases, the embodiment can provide a point of the focus for interaction. For some applications this can be more important than showing subtle details and changes in facial expressions and gestures.

4. Emotional behavior and social relationships

We discuss three topics in order to illustrate how research on ECAs can incorporate issues that deal with emotion and affect and the development of social relationships between humans and embodied agents acting in ambient intelligence environments. Computational modeling of emotions is one of them; the development of friendship relations is another. Both topics are receiving considerable attention, although applications are hard to find. The third topic we want to mention is humor. This is a rather undeveloped area from a computational point of view. However, its importance in natural interaction should be clear. That is, as Cowie [16] mentioned "A useful way of making the point is in terms of artificial agents. If they are going to show emotion, we surely hope that they would show a little humor too."

4.1. Emotions and affect

Facial expressions, body posture and intonation in speech are the main modalities to express nonverbal emotion. Human beings do not express emotions using one modality only. Generally, they have their emotions displayed using a combination of modalities that interact with each other. We cannot consider one modality in isolation. Facial expressions are combined with speech. There are not only audio or visual stimuli, but also audio–visual stimuli when expressing emotions. A smile gesture will change voice quality; variations in speech intensity will change facial expression, etc. Attitude, mood and personality are other factors that make interpretation and generation of emotional expressions even less straightforward. In addition we can have different intensities of emotion and the blending of different emotions in an emotional expression.

In embodied agents we should consider combinations and integration of speech, facial expressions, gestures, postures and bodily actions. It should be understood that these are displays and that they should follow from some emotional state that has been computed from sensory inputs of a human interactant, but also from an appraisal of the events that happen or have happened simultaneously or recently. A usual standpoint is that of appraisal theory, the evaluation of situations and categorizing arising affective states. It should be understood that what exactly is said and what exactly is done in a social and emotional setting is not part of the observations above. The importance of the meaning of words, phrases and sentences, uttered and to be interpreted in a specific context is not to be diminished.

An example of an embodied agent that has been designed to show affect in its interaction with students is presented in Fig. 3.

In the educational environment that is displayed here we have a student using a haptic device (represented in the virtual world as an injection needle), a virtual patient on the left monitor and a virtual tutor displayed on the right monitor. There is interaction between student and patient and between tutor and student. The tutor monitors the student's progress and uses this information to make assumptions about the student's emotional state. This makes it possible to express affect through voice and facial expression, making the process more effective and motivating [17].

Fig. 3. Virtual embodied agents in an educational environment.

4.2. Friendship

One of the issues we investigated was how aspects of personal attraction or friendship development can be made part of the design of an embodied agent that is meant to provide an information service to a human partner. As a 'lay psychologist', we all know that people that you like (or your friends) are able to help you better, teach you better, and generally are more fun to interact with, than people that you do not like. However, 'liking' is person dependent. Not everybody likes the same person, and one person is not liked by everyone. These observations sparked our interest in the application, effects and design of a 'virtual friend'. An agent that observes its user, and adapts its personality, appearance and behavior according to the (implicit) likes and dislikes of the user, in order to 'become friends' with the user and create an affective interpersonal relationship. This agent might have additional benefits over a 'normal' ECA in areas such as teaching, navigation assistance and entertainment.

There is extensive knowledge about human interpersonal relationships in the field of personality and social psychology. Aspects of friendship that need to be considered in embodied agent design are gender (e.g., activity-based men's friendship vs. affectively based women's friendship), age, social class and ethnic background. Effects of friendship on interaction include increase of altruistic behavior, a positive impact on task performance and an increase in self-disclosure. Interpersonal attraction is an important factor in friendship. It is governed by positive reinforcements, and similarity between subjects is a key factor. Similarity of attitudes, personality, ethnicity, social class, humor, etc., reinforces the friendship relationship. Other issues are physical attractiveness (the 'halo effect') and reciprocity of liking (whether we think that the other person likes us). In [18], we discussed the translation of the main aspects of human–human friendship to human–ECA friendship and how we can incorporate this translation in the design process of an ECA, using a scenario-based design. One observation is that it is important to distinguish between the initial design of an ECA and the possibility to change the ECA characteristics according to an adaptation strategy based on knowledge obtained by interacting with a particular user.

Trust is not only important in friendship relations. Bickmore [19] introduced relational agents, embodied agents that attempt to build a trusting, empathetic relationship with their conversational partners. Interaction with embodied agents that play helping roles like a teacher, a doctor or a personal assistant (virtual butler) can be made more effective and comfortable when such a relationship can develop over time.

4.3. Humor

Humans use humor to ease communication problems and in a similar way humor can be used to solve communication problems that arise with human–computer interaction. For example, humor can help to make the imperfections of natural language interfaces more acceptable for the users and when humor is sparingly and carefully used it can make natural language interfaces much friendlier. During these years the potential role of ECAs was not at all clear, and no attention was paid to their possible role in the interface.

Humans employ a wide range of humor in conversations. Humor support, or the reaction to humor, is an important aspect of personal interaction and the given support shows the understanding and appreciation of humor. There are many different support strategies. Which strategy can be used in a certain situation is mainly determined by the context of the humorous event. The strategy can include smiles and laughter, the contribution of more humor, echoing the humor and offering sympathy. In order to give full humor support, humor has to be recognized, understood and appreciated. These factors determine our level of agreement on a humorous event and how we want to support the humor.

Humor plays also an important role in interpersonal interactions. From the many SRCT experiments [4,20], we may extrapolate that humor will play a similar role in human–computer interactions. This has been confirmed with some specially designed experiments. There is not yet much research going on into embodied agents that interpret or generate humor in the interface. In [21], we discussed how useful it can be, both from the point of view of humor research and from the point of view of ECA research, to pay attention to the role of humor in the interaction between humans and the possibility to translate it to the interactions between humans and ECAs.

Graphics, animation and speech synthesis technology make it possible to have embodied agents that can display smiles, laughs and other signs of appreciation of the interaction or explicitly presented or generated humor. There are many applications that can profit from being able to employ such embodied agents. The designer of the interface can decide when in certain scenarios of interaction agents should display such behavior. However, much more in the line of research on autonomous (intelligent and emotional) agents we rather have an agent understand why the events that take place generate enjoyment by its conversational partner and why it should display enjoyment because of its appreciation of a humorous situation.

5. Conclusions and future research

Ambient intelligence is said to consist of ubiquitous computing + social and intelligent user interfaces allowing social interaction. This also assumes that in ambient intelligence 'the real world is the interface'. Presently, in (traditional) human–computer interaction more and more applications assume that interaction should be socially formed. In this paper, it is argued that this will be even more the case in smart and ambient intelligence environments. Embodied agents allow the development of affinitive relationships with their human partners and can therefore help to fulfill the need of affiliation in ambient intelligence environments. However, introducing these virtual humans in our environments requires display and rendering possibilities, embedded in the environment. It also requires the modeling of multi-party interaction, where the members of the party can be virtual and can be real humans and they are not necessarily aware of the fact that their conversational partner is human or virtual.

Most of the research in ambient intelligence deals with the question how the environment is able to identify and model users' activities, rather than how the user will be able to communicate with the environment. In more traditional environments multi-modality in interactions has received attention [2], but it has hardly been investigated how these results can be used in environments rather than in situations where the user uses keyboard, mouse and screen. Moreover, most of the research on ambient intelligence does not take into account that maybe people will get lost in ambient intelligence, do not know who to 'talk' to and will not be able to build some kind of social relationship with the anonymous environment that nevertheless is assumed to support them, observe them and keep track of their activities. Rather than introducing intelligent remote controls to control the environment we give preference to embodied agents that are there when they are needed and that can explain alternatives in choices, provide suggestions about actions to take and, when they are sufficiently aware of preferences and possibilities, take actions themselves.

References

[1] Weiser M. The computer for the 21st century. Scientific American 1991;265(3):94–104.
[2] Nijholt A. Multimodal interactions in ambient intelligence. In: Verhaegh W, Korst J, Aarts E, editors. Algorithms in ambient intelligence. Philips Research Book Series. Dordrecht: Kluwer Academic Publishers; 2003 [Chapter 2].
[3] Montore G, Alaman X, Haya PA. A plug and play spoken dialogue interface for smart environments. In: Gelbukh A, editor. Computational linguistics and intelligent text processing. Lecture notes in Computer Science, vol. 2945. Berlin: Springer; 2004.
[4] Reeves B, Nass C. The media equation. Cambridge: Cambridge University Press; 1996.
[5] Gosling SD, Ko Sei Jin, Mannarelli T, Morris ME. A room with a cue: personality judgments based on offices and bedrooms. Journal of Personality and Social Psychology 2002;82(3):379–98.
[6] Shechtman N, Horowitz LM. Media inequality in conversation: how people behave differently when interacting with computers and people. In: SIGCHI-ACM CHI 2003: New Horizons. New York: ACM; 2003. p. 281–8.
[7] Goffman E. Behavior in public spaces. New York, USA: The Free Press, Simon & Schuster Inc.; 1963.
[8] Pertaub D-P, Slater M, Barker C. An experiment on public speaking anxiety in response to three different types of virtual audience. Presence: Teleoperators and Virtual Environments 2002;11(1):68–78.
[9] Nijholt A. Agents, believability and embodiment in advanced learning environments. In: Okamto T, Hartley R, Kinshuk, Klus JP, editors. Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT 2001), Madison, Wisconsin, August 2001. Los Alamitos, CA: IEEE Computer Society; 2001. p. 457–9.
[10] Lester JC, et al. The persona effect: affective impact of animated pedagogical agents. CHI '97 Human Factors in Computing Systems. New York: ACM; 1997. p. 359–366.
[11] Cassell J, Sullivan J, Prevost S, Churchill E, editors. Embodied conversational agents. Cambridge: The MIT Press; 2000.
[12] Argyle M, Cook M. Gaze and mutual gaze. Cambridge: Cambridge University Press; 1976.
[13] Torres O, Cassell J, Prevost S. Modeling gaze behavior as a function of discourse structure. First International Workshop on Human Computer Conversations, Bellagio, Italy, 1979.
[14] Heylen D, van Es I, van Dijk B, Nijholt A. Experimenting with the gaze of a conversational agent. In: van Kuppevelt J, Dybkjaer L, Bernsen NO, editors. Natural, intelligent and effective interaction in multimodal dialogue systems. Dordrecht: Kluwer Academic Publishers; 2004.
[15] Kendon A. Gesticulation and speech: two aspects of the process of utterance. In: Key MR, editor. The relation of verbal and nonverbal communication. The Hague, The Netherlands: Mouton; 1980.
[16] Cowie R. Describing the emotional states expressed in speech. Proceedings of the ISCA Workshop on Speech and Emotion, Belfast, September 2000. p. 11–8.
[17] Nijholt A, et al. Multimodal integration of haptics, speech and affect in an educational environment. Internal Report, University of Twente, February 2004.
[18] Stronks B, Nijholt A, van der Vet P, Heylen D. Designing for friendship: becoming friends with your ECA. In: Marriott A, Pelachaud C, Rist T, Ruttkay Zs, editors. Proceedings of the Embodied Conversational Agents—Let's Specify and Evaluate Them! Bologna, Italy, 2002. p. 91–7.
[19] Bickmore TW. Relational agents: effecting change through human–computer relationships. PhD thesis, Massachusetts Institute of Technology, February 2003.
[20] Morkes J, Kernal HK, Nass C. Effects of humor in task-oriented human–computer interactions and computer-mediated communication: a direct test of SRCT theory. Human–Computer Interaction 1999;14:395–435.
[21] Nijholt A. Embodied agents: a new impetus to humor research. In: Stock O, Strapparava C, Nijholt A, editors. The April fools day workshop on computational humour. Proceedings of the Twente Workshop on Language Technology 20 (TWLT 20), Trento, Italy, 2002. p. 101–11.