# Disappearing Computers, Social Actors and Embodied Agents

Anton Nijholt

*Department of Computer Science,*
*University of Twente, Enschede, The Netherlands.*
*anijholt@cs.utwente.nl*

## Abstract

*Presently, there are user interfaces that allow multimodal interactions. Many existing research and prototype systems introduced embodied agents, assuming that they allow a more natural conversation or dialogue between user and computer. Here we will first take a look at how in general people react to computers. We will look at some of the theories, in particular the CASA ("Computers Are Social Actors") paradigm, and then discuss how new technology, for example ambient intelligence technology, needs to anticipate the need of humans to build up social relationships. One way to anticipate is to do research in the area of social psychology, to translate findings there to the human-computer situation and to investigate technological possibilities to include human-human communication characteristics in the interface. For that reason we will discuss embodied conversational agents, the role they can play in human-computer interaction (in face-to-face conversation), in ambient intelligence environments and in virtual communities.*

## 1. Introduction

In the "Media Equation" Byron Reeves and Clifford Nass report about their experiments on human-computer interaction where humans assign human characteristics to computers [14]. Many experiments have been done after this book has been published. They 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 positive and politely than when answering similar questions posed by the second computer. Clearly, people don't like to offend a computer that has 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 very much 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 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."*

For a future situation where a house, a sitting room, a working room, an office and in fact every environment and its objects allow perception of what is going on in the environment and allow interaction by its occupants and visitors to exchange information (including emotions), it is certainly useful to investigate how we can design social interfaces, emphasizing human-to-human communication properties, rather than concentrating purely on designing intelligence.

These ambient intelligence environments are intimately integrated with our everyday environments. We should feel comfortable within them, 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
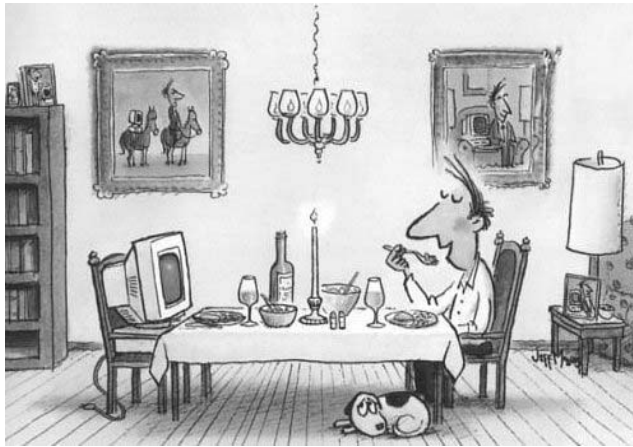
**Figure 1. The "Media Equation".**

need support in our activities. The question that needs to be addressed is which aspects of human-human interaction we don't want to loose when we will spend much of our future time in these environments and also spend much time interacting – maybe not always explicitly – with these environments.

One important issue in human-computer interaction is the design of the interface. Where is the interface when, as has been observed, ambient systems become one with physical settings, that is, where 'the real world is the interface'? When we have a question or need other support, who or what do we address in these environments? Will we not feel lost in ambient intelligence?

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 embodied conversational agents in order to be able to maintain desirable characteristics of human-human communication, e.g., to establish a short- or long-term relationship. Embodied agents can offer intelligence and emotion and therefore communication properties that help to make us feel being appreciated and that make us feel being understood. It makes it possible for us to act in a smart, but also in a social environment.

## 2. Social actors, interpersonal relationships and the disappearing computer

In the previous section we introduced the computer as a social actor. In human-computer interaction we recognize characteristics of human-human interaction. There is human-like behavior when interacting with the computer and human-like behavior of the computer is expected. Can we expect similar behavior when the user

is interacting with an environment rather than with a desktop screen? In future environments computers will be embedded in walls, furniture, cloths, and in objects that are natural in the environment. Moreover, there is communication between these embedded computational devices allowing a much more comprehensive overview of environment and events taking place than is possible with a single computing device.

How will humans interact with such environments? Are they able to build some kind of relationship with these environments like they are able to build relationships with a computer that is perceived as a social actor? Or do we need to introduce explicit social actors, that is, embodied conversational agents, in these environments with which users can communicate and exchange information in intelligent and social ways in order to fulfil 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 more and more resemble each other. 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 especially 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 an 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 leave their traces and their personalities show in their (personal) 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.

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 the partner's satisfaction when arranging things. Sometimes arranging itself is entertaining. Both interaction and information exchange can be goals themselves, e.g., when we enter conversations with our

children or colleagues. Efficiency is not necessarily the starting point when engaging in these conversations.

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 [15]. 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 toward 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 much more communion and agency relationship statements, used more words and spent more time in conversation.

Not all modalities that can be employed in human-computer interaction lend themselves to the same degree to the different types of performance that we distinguished above. In human-human interaction nonverbal cues play an important part in the relationship track of communication. Hence, we can ask whether we can recognize and interpret these communication aspects in human-computer interaction and whether they can play a similar role. From the SRCT perspective we know that humans react socially on social computer behavior and having the computer display more cues about its social behavior may strengthen the social reaction.

Obviously, 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. Nevertheless, there will be many situations, especially in a personal environment, where people will prefer communicating with systems that show knowledge of the user and display 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. Nonverbal cues in human-human interaction need to be represented in human-computer interaction and in interacting in ambient intelligence environments in order to obtain the same possibilities for reaching the different conversation goals that can be achieved in human-human interaction. An obvious way to allow these nonverbal cues to happen is the use of embodied agents.

## 3. Embodied conversational agents and nonverbal behavior

Embodied conversational agents (ECAs) have become a well-established research area. 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. These agents are used to inform and explain or even 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. [8] showed that a display of involvement by an embodied conversational agent motivates a student in doing (and continuing) his or her learning task. Some examples of embodied conversational agents are shown in Figure 2. From left to right we see: Jennifer James, a car saleswoman who attempts to build relationships of affection, trust and loyalty with her customers, Karin, informing about theatre performances and selling tickets,
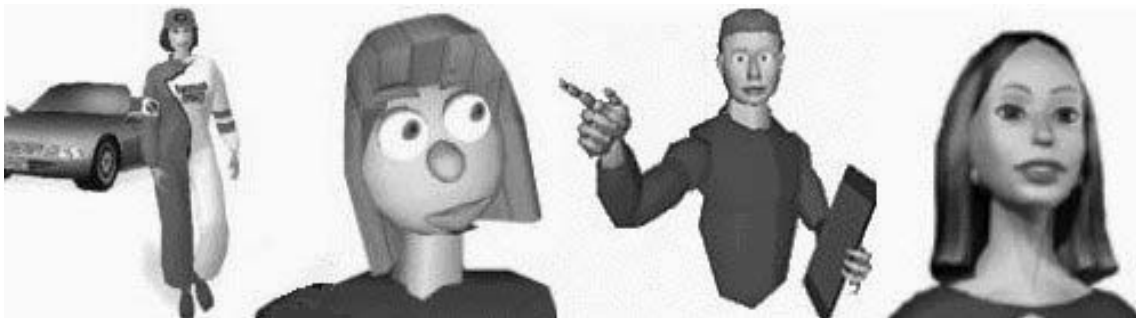


**Figure 2. Examples of 2D and 3D embodied agents.**

Steve, educating a student about maintaining complex machinery, and Linda, a learning guide.

Embodiment allows more multimodality, 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. [2] for a collection of chapters on properties and impact of embodied conversational agents (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.

### 3.1. 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, 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.

### 3.2. 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 [1,17]. 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 [6], we investigated the effects of different styles of gaze of Karin on the conversation. We had forty-eight 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.

### 3.3. 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 [7] 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 tot 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. When, for example, we introduce a guide in our virtual environments this is a main issue and probably more important than detailed facial expressions and gestures.

## 4. Emotional behavior and social relationships

We discuss three topics in order to illustrate how research on embodied conversational agents 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 an other. 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 mentioned Cowie [3] "*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 humour too.*"

### 4.1. Emotions and affect

Facial expressions and speech are the main modalities to express nonverbal emotion. Human beings do not express emotions using facial expressions and speech 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.

In Figure 3 we display Cyberella, an embodied agent, developed at DFKI in Saarbrücken. This agent is working as a receptionist. For example, she can provide directions to the office of a staff member. Since she has been provided with an affective model, she also reacts emotionally to a visitor's utterances when appropriate [4].

### 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 don't 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 it's user, and adapts it's 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



**Figure 3. Cyberella, a virtual receptionist.**

'normal' embodied conversational agent 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 ECA 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 [16] 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.

### 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 embodied conversational agents 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 [9] 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 [12] we discuss how useful it can be, both from the point of view of humor research and from the point of view of embodied conversational agent 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 embodied conversational agents.

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 an humorous situation.

## 5. Ambient intelligence and embodiment: Conclusions

Most of the research in ambient intelligence deals with how the environment is able to identify and model users' activities, rather than how the user is able to communicate with the environment. In more traditional environments multimodality in interactions has received attention [10,11,13], but it has hardly been investigated how these results can be used in environments rather than in situations where the user explicitly addresses a computer screen. Moreover, most of the research on ambient intelligence does not take into account that maybe people will get lost in ambient intelligence, don't know who to 'talk' to and will not be able to build some kind of social relationship with the anonymous environment that nevertheless supports them, observes them and keeps track of their activities.

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 assumed that this will be even more needed in 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.

## References

[1] M. Argyle and M. Cook. *Gaze and Mutual gaze*. Cambridge University Press, 1976.

[2] J. Cassell, J. Sullivan, S. Prevost and E, Churchill (eds.). *Embodied Conversational Agents*. The MIT Press, 2000.

[3] R. Cowie. Describing the emotional states expressed in speech. Proceedings of the ISCA Workshop on *Speech and Emotion*. Belfast, September 2000, 11-18.

[4] P. Gebhard. Enhanching Embodied intelligent agents with affective user modelling. *UM2001, 8th International Conferenc*e, J. Vassileva and P. Gmytrasiewicz, (eds.), Berlin, Springer, 2001.

[5] S.D. Gosling, Sei Jin Ko, T. Mannarelli and M.E. Morris. A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology* 82 (2002), No. 3, 379-398.

[6] D. Heylen, I. van Es, B. van Dijk & A. Nijholt. Experimenting with the Gaze of a Conversational Agent. Chapter in *Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems.* J. van Kuppevelt, L. Dybkjaer & N.O. Bernsen (eds.), Kluwer Academic Publishers, 2003.

[7] A. Kendon. Gesticulation and speech: two aspects of the process of utterance. In: *The relation of verbal and nonverbal communication*. M.R. Key (ed.), Mouton, The Hague, the Netherlands, 1980.

[8] J.C. Lester et al. The persona effect: Affective impact of animated pedagogical agents. *CHI '97 Human Factors in Computing Systems*, ACM, 1997, 359-356.

[9] J. Morkes, H.K. Kernal and C. Nass. Effects of humor in task-oriented human-computer interactions and computer-mediated communication: a direct test of SRCT theory. *Human-Computer Interaction*, 14, 1999, 395-435.

[10] A. Nijholt. From virtual environment to virtual community. In: *New Frontiers in Artificial Intelligence*. Joint Terano, T., Nishida, T., Namatame, A., Tsumoto, S., Ohsawa, Y., Washio, T., (Eds.). Lecture Notes in Computer Science. Vol. 2253, Springer, Tokio, 2001, 19-26.

[11] A. Nijholt. Multimodal Interactions in ambient intelligence. Chapter in *Algorithms in Ambient Intelligence*. Philips Research Book Series, Kluwer Academic Publishers. W. Verhaegh, J. Korst & E. Aarts (eds.), 2003.

[12] A. Nijholt. Embodied Agents: A New Impetus to Humor Research. *The April Fools Day Workshop on Computational Humour*, O. Stock, C. Strapparava & A. Nijholt (eds.), In: Proc. Twente Workshop on Language Technology 20 (TWLT 20), Trento, Italy, 2002, 101-111.

[13] A. Nijholt and D. Heylen. Multimodal communication in inhabited virtual environments. *International Journal of Speech Technology*, Vol. 5, Issue 4, November 2002, 343-354.

[14] B. Reeves and C. Nass. *The Media Equation*. Cambridge University Press, Cambridge, 1996.

[15] N. Shechtman and L.M. Horowitz. Media inequality in conversation: how people behabe differently when interacting with computers and people. In: *SIGCHI-ACM CHI 2003: New Horizons*, ACM, New York, 2003, 281-288.

[16] B. Stronks, A. Nijholt, P. van der Vet and D. Heylen (2002). Designing for friendship: Becoming friends with your ECA. In: Proc. *Embodied conversational agents - let's specify and evaluate them!* A. Marriott, C. Pelachaud, T. Rist and Zs. Ruttkay (eds.), Bologna, Italy, 91-97.

[17] O. Torres, J. Cassell and S. Prevost. Modeling gaze behavior as a function of discourse structure. First International Workshop on *Human Computer Conversations*. Bellagio, Italy, 1979.