Unexploited Dimensions of Virtual Humans

Zsófia Ruttkay, Dennis Reidsma, Anton Nijholt
Human-Media Interaction, Department of Computer Science,
University of Twente, Enschede, the Netherlands
{zsofi,dennisr,anijholt}@cs.utwente.nl

Abstract

Virtual Humans are on the border of fiction and realism: while it is obvious that they do not exist in reality and function on different principles than real people, they have been endowed with human features such as being emotionally sensitive. In this article we argue that many dimensions, both human-like ones and ones made possible by the computer technology, are still unexploited to increase the effectivity and engagement of interaction with VHs.

1 Introduction

The concept of *Human Computing* unifies several objectives: to make the usage of computers easy and natural, allowing such very 'human' behaviors like being emotional or bored, endow computer systems with adaptive and empathic response, facilitating applications where joy and engagement become more important than the problem-solving oriented 'categorical computing' practice.

To achieve this, application interfaces need to be able to handle more subtle interactions than those afforded by more basic "push-button type" interfaces. Very sophisticated sensing modules are needed to cope with subtleties in the behavior of the user. Also, the application needs to be able to communicate information to the user with a comparable level of complexity.

Virtual Humans (VHs) [Plantec, 2004] – also known under other names like humanoids [Thorisson, 1996], embodied conversational agents [Cassell et al., 2000] or intelligent virtual agents – are computer models resembling humans in their bodily look and their communication capabilities. In this paper we focus on their role in human-computer interaction, assuming a scenario where one or more VHs interact with one real person, in order to accomplish some task. We keep the traditional terms *user* and *task*, but will use it in a broadened sense. Because of their human-likeness, VHs seem to be the ultimate interface in the Human Computing age. From the point of view of HCI, there are two major motivation of 'putting a VH on the screen':

- By replacing the traditional computer-specific interaction modalities (keyboard, mouse) by the natural communicational capabilities of people, the services of computers will be *accessible to a broad population*, without respect to (computer) literacy, cultural and social background;
- VHs make *new applications* possible, where they fulfill traditional human roles, such as tutor, salesperson, and partner to play a game or chat with. Moreover, there are applications where the VH is in an entirely new role, without parallel in real life, such as human-like characters with fictional capabilities in (educational) games [Gustafson et al., 2004], interactive drama [Mateas and Stern, 2003] or Virtual Reality applications [Abaci et al., 2004], or a VH taking partial or split role known from real-life situations [Baylor and Ebbers, 2003].

Often it is not easy to separate the two aspects in an application: e.g. a VH explaining how to operate a device [Badler, 2002], can be seen as a friendly interface replacing queries and search in an on-line textual help, but it also extends the services, by being able to demonstrate operations, possibly in a way tailored to the user's capabilities (e.g. handedness).

The potentials of VHs are huge. Challenges to current state of the art of separate disciplines – e.g. speech recognition and language understanding, or computer vision – , have been identified as a must to improve technical quality, and related to this, engagement and effectivity of VHs [Gratch et al., 2002]. Also, the necessity of cooperation of different disciplines, and particularly, dedicated studies providing basis for computational models of human-human interaction, have been underlined [Isbister and Doyle, 2004].

In this paper, we look at VHs from the Human Computing perspective, and dwell on possible improvements of VHs of a different nature. Namely, we outline features which have not (or hardly) been exploited, and relate the virtual and the human domain. Several of these are technically feasible, or would require some routine engineering development. For instance, current VHs look as puppets took out of a box, new forever and unchanged, very much unlike real people, who may change dress every day, look more tired at the end of a busy day than in the morning, and their mood and mental state is not exactly the same every morning. This paper is a 'twin' of another recent paper by us [Ruttkay et al. 2006a], where we concentrated among other things on the practical

value of some subtle aspects of human communication which have often been left out of the repertoire of 'machine humans' since they are seen as 'incorrect' or 'undesirable' in some way, such as disfluency and ambiguity. Here we look at the *fictional/real* polemic: a VH, on the one hand, is a fictional character, created and empowered by computer technology, so there is no need to model certain human characteristics, especially those which 'make no sense' or mean limitations for a fictional character, such as getting tired. Hence it is clear, and from the fictional point of view natural, that a VH would never ever get exhausted physically or mentally as real people do. On the other hand, if they remain ever fit, they are a mismatch to the human partner, who does get tired after some time. Such a mismatch could not only produce a feeling of inferiority and thus discomfort by the user, but may put him to danger, if he would try to keep up with the constant high performance of e.g. a virtual fitness trainer.

We see two reasons why the above and other similar aspects of VHs have not been considered:

- There is uncertainty about how *human-like* [Koda and Maes, 1996] and *realistic* [Dautenhahn, 2004] VHs should be. Is it at all wanted that a VH, who is not made of flesh and blood, gets tired, or has an own wardrobe that also reflects the seasons?
- Besides the hard core issues of e.g. understanding language, subtle features like style may be considered as not necessary, or at least not of high priority to be dealt with.

In the next section, we introduce three novel applications being developed at our group: the Virtual Dancer, the Virtual Conductor, and the Virtual Trainer. In Section 4 we discuss one by one the yet unexplored aspects of VHs and explain their potential merits, illustrating them with examples from these applications. Finally, in the Discussion we return to the question of fiction and reality of VHs.

3 VHs in novel applications

In this section, we present three applications currently being developed at the HMI (Human Media Interaction) research group: the Virtual Dancer [Reidsma et al., 2006], the Virtual Conductor [Bos et al., 2006], and the Virtual Trainer [Ruttkay et al., 2006b]. These three applications are shown in Figure 1. These seemingly very different applications share some basic features, and have actually been developed relying on a similar framework. In all three applications, the VH:

- has visual and/or acoustic perception capabilities,
- has to monitor and react to the user continuously,
- has to use subtle variants of a motion repertoire generated on the fly, and
- uses both acoustic (music, speech) and nonverbal modalities in a balanced and strongly interwoven manner.

3.1 The Virtual Dancer

In a recent application built at HMI, a virtual human – the Virtual Dancer – invites a real partner to dance with her [Reidsma et al., 2006] (see Figure 2). The Virtual Dancer dances together with a human 'user', aligning its motion to the beat in the music input and responding to whatever the human user is doing. The system observes the movements of the human partner by using a dance pad to register feet activity and the computer vision system to gain information about arm and body movements. Using several robust processors, the system extracts global characteristics about the movements of the human dancer like how much (s)he moves around or how much (s)he waves with the arms. Such characteristics can then be used to select moves from the database that are in some way 'appropriate' to the dancing style of the human dancer.

There is a (non-deterministic) mapping from the characteristics of the observed dance moves to desirable dance moves of the Virtual Dancer. The interaction model reflects the intelligence of the Virtual Dancer. By alternating patterns of following the user or taking the lead with new types of dance moves, the system attempts to achieve a mutual dancing interaction where both human and virtual dancer influence each other. Finding the appropriate nonverbal interaction patterns that allow us to have a system that establishes rapport with its visitors is one of the longer term issues being addressed in this research.
Clearly, the domain of dancing is interesting for animation technology. We focus on the interaction between human and virtual dancer. The interaction needs to be engaging, that is,

Figure 1. Three novel applications: Virtual Dancer, Virtual Conductor, and Virtual Trainer

interesting and entertaining. First experiences with demonstration setups at exhibitions indicate that people are certainly willing to react to the Virtual Dancer (see Figure 1).

3.2 A Virtual Conductor

We have designed and implemented a virtual conductor [Bos et al., 2006] that is capable of leading, and reacting to, live musicians in real time. The conductor possesses knowledge of the music to be conducted, and it is able to translate this knowledge to gestures and to produce these gestures. The conductor extracts features from the music and reacts to them, based on information of the knowledge of the score. The reactions are tailored to elicit the desired response from the musicians.

Clearly, if an ensemble is playing too slow or too fast, a (human) conductor should lead them back to the correct tempo. She can choose to lead strictly or more leniently, but completely ignoring the musicians' tempo and conducting like a metronome set at the right tempo will not work. A conductor must incorporate some sense of the actual tempo at which the musicians play in her conducting, or else she will lose control. If the musicians play too slowly, the virtual conductor will conduct a little bit faster than they are playing. When the musicians follow him, he will conduct faster yet, till the correct tempo is reached again.

The input of the virtual conductor consists of the audio from the human musicians. From this input volume and tempo are detected. These features are evaluated against the original score (currently stored in MIDI) to determine the conducting style (lead, follow, dynamic indications, required corrective feedback to musicians, etc) and then the appropriate conducting movements of the virtual conductor are generated. A first informal evaluation showed that the Virtual Conductor is capable of leading musicians through tempo changes and of correcting tempo mistakes from the musicians. Computer vision has not yet been added to the system. That is, musicians can only interact with the conductor through their music. In a future implementation we can look at the possibility to have the conducting behavior directed to (the location of) one or more particular instruments and their players.

3.3 The Virtual Trainer

The *Virtual Trainer* (VT) application framework is currently under development [Ruttkay et al., 2006a] and involves a virtual human on a PC, who presents physical exercises that are to be performed by a user, monitors the user's performance, and provides feedback accordingly at different levels. Hence, our VT should fulfill most of the functions of a real trainer: it not only demonstrates the exercises to be followed, it should also provide professionally and psychologically sound, human-like coaching. Depending on the motivation and the application context, the exercises may be general fitness exercises that improve the user's physical condition, special exercises to be performed from time to time during work to prevent for example RSI (Repetitive Strain Injury), or physiotherapy exercises with medical indications. The focus is on the reactivity of the VT, manifested in natural language comments relating to readjusting the tempo, pointing out mistakes or rescheduling the exercises. When choosing how to react, the static and dynamic characteristics of the user and the objectives to be achieved are to be taken into account and evaluated with respect to biomechanical knowledge and psychological considerations of real experts. For example, if the user is just slowing down, the VT will urge him in a friendly way to keep up with the tempo, acknowledge with cheerful feedback good performance and engage in a small talk every now and then to keep the user motivated.

The VT is adaptable and adaptive in several respects. The embodiment can be chosen such that it reflects the geometrical and physiological motion characteristics of the user. For this purpose, some data (age, gender, weight, goal of the training) may be asked for, or gained by computer vision. The motion characteristics may be gained by in an initial calibration session by computer vision, analyzing the user's motion perform a few special moves. The exercises to be presented may be authored by an authorized person, such as a real physiotherapist to whom the VT acts as an 'assistant'. The motion and exercise repertoire of the VT may be extended, by providing an exercise editing interface, allowing also the incorporation of complex motions which were preacted and motion captured.

The VT keeps record of the sessions with the user, and interprets his performance in the light of short and long-term history. Also the VT addresses the user in a personal man-

Figure 2. Interacting with the Virtual Dancer

ner, using his name and a style most appropriate for the given user's age.

4 Unexploited aspects of VHs

In the past decade, much effort has been spent on improving the human-likeness of individual modalities of VHs, such as improving the quality of synthesized speech [Van Moppes 2002], modeling expressive gesturing of humans [Hartmann et al., 2005], deriving computational models to capture the kinematics [Wachsmuth and Kopp, 2002], providing means to fine-tune the effort and shape characteristics of facial expressions and hand gestures [Chi et al., 2000], model gaze and head behavior, add biological motions like blinking or idle body motion [Egges et al., 2004].

The fusion of multiple modalities has been dealt with, from the point of view of timing of the generated behavior, and the added value of using multiple modalities in a redundant way. It has been suggested that VHs, just as real people, should be endowed with a style, reflected in the usage of verbal and non-verbal modalities to express some meaning, and the intonational and motion characteristics of the single modal signals [Ruttkay et al., to appear]. Besides the features typical for the VH as an individual, his (assumed) social, cultural and professional background should be the components which contribute to the style. The importance of cultural and social connotation of a VH has been pointed out [Payr and Trappl, 2004, Prendinger and Ishizuka, 2001].

Modeling emotions, mood and personality [Gratch and Marsella, 2001] and their benefits in judging the VHs have been extensively addressed. Initially, the 6 basic emotions were to be shown on the face [Ekman, 1989], which has been followed by research on taking into account display rules, resulting in emotions to be hidden [Poggi et al., 2001], or overcast by fake expressions e.g. to hide lies [Rehm and Andre, 2005], studying principles to show mixed emotions on the face, to reflect emotions in gesturing [Noot and Ruttkay, 2004].

In addition to emotions, the importance of small talk in building a common ground and trusting the VH [Bickmore and Cassell, 2000], as well as back-channeling have been emphasized. Having long-term attitude towards a VH, like friendship, has been pointed out [Stronks et al., 2002].

4.1 Embodiment: beyond the perfect and generic

VHs should go beyond the present state of generic, doll-like and usually perfect-looking (symmetrical, young and spotless face and body) design. To begin with, the embodiment should reflect age, ethnicity and social status most appropriate for the given application and the user group. Besides, a VH should have individual features, which may make him easier to identify, remember and enjoy.

The individuality may be in granularity of detail of photorealistic nature (e.g. applying subtle texture for the skin), or in exaggerated cartoonish features [Liang et al., 2002] or variety resulted from 'noisy' parameters [DiPaola, 1991]. The phenomena 'uncanny valley', formulated by Mori [Mori, 1982] originally for robots stating that increasing realism, after a certain degree, dramatically decreases the perceived human quality of the robot, is widely assumed to apply for VHs too.

The technology allows that the embodiment (and also, the communicational and mental characteristics) of a VH should be chosen according to the given user. The 'mirroring phenomena', stating that people are positively prejudiced to others who resemble, by and large, themselves, could be turned to good use: by 'looking' at the user first, the best matching VH embodiment could be chosen, based on the assumption that such a VH will be the most trusted and effective [Bailenson, and Yee, 2005]. We believe that there is much use for this type of 'adaptive appearance' for at least the Virtual Dancer and the Virtual Trainer, certainly concerning the gender and age parameters. For the VT, the body geometry in general or based on measurements could be a useful parameter to be adjusted. A recent study on preference of the virtual eHealth advisors points in the direction that a somewhat bulky figure is more appreciated than one with the ideal weight.

Another source of variety is in subtle temporal differences of the appearance of a given VH. Changes in outfit (hair, clothing) and signs of physical state would make the VH more enjoyable and life-like. Moreover, such variations in appearance could be used to reinforce the presence in the geographical location and time of the user (see below). How about having the VT pop up on a hot day in appropriate summer dress, sun-burnt? And would not it be nice if the VT would start sweating too after some time, not only the user? These temporal changes should be consistent with each other and with the identity and history of the VT (see below).

4.2 VH- with an own history and identity

"Who are you?" do people ask (usually as one of the first questions) from their VH interlocutor. The answer is a name, may be extended with the services the VH can offer. In case of chat bots, a date of birth may be given, and the creator may be named as 'father', such as in the case of Cybelle [AgentLand, 2006]. Notably, the date of 'creation' makes sense in the fictional framework only. Moreover, any inquiry about further family members is not understood. The personal history is similarly shallow and inconsistent as of her hobbies: she has a favorite author, but cannot name any title by him. Deviations from this common solution can be found, when the VH is to stand for a real, dead person [Bernsen et al., 2004], and the very application is to introduce the reincarnated real person and his history to the user. The other extreme is feasible when the VH is in a role like a museum guide [Kopp et al., 2003], where his refusal 'to talk about any personal matters' sounds to be a natural reaction. But in other applications, where it would be appropriate, we would never know about the family, schooling, living conditions, acquaintances and other experiences of the VH, neither about his favorite food or hobbies. One may argue that that is enough, or even preferred, to remain 'to the point' in well-defined task-oriented application like a weather re-

porter or trainer. However, even in such cases in real life some well-placed reference to the expert's 'own life and identity' breaks the business-like monotonicy of the service, and can contribute to create common ground and build up trust. B. Hayes-Roth endowed her Extempo characters with some own history as part of their 'anima' [Hayes-Roth and Doyle, 1998]. From the recent past, we recall a Dutch weather forecast TV reporter who added, when a certain never heard-of Polish town was mentioned as the coldest place in Europe, that this town is special for him as his father was born there. But he could have noted about some other aspects like special food or customs he experienced or knows of from that place. In case of a real fitness trainer's video, it is remarkable how the task-related talk is interwoven with references to the presenter's personal experience on where she learnt the exercises, what she found difficult, etc. A VH could use his personal background to generate just some 'noise-like small talk' in addition to the task-related conversation, or to relate it to the stage of task completion or difficulty and the reactions from the user, in order to increase the user's commitment. So for instance, a VT may include not task-related small talk at the beginning or during resting times, or add task-related background information to keep the user motivated during a long and/or difficult exercise.

In order to make a VH 'personal', it is not enough to endow him with a 'personal history'. Some mechanisms should be provided to be able to decide when and what piece of personal information to tell. E.g. to derive if there is something in the personal knowledge of the VH which could be related to the factual, task-oriented information to be told. This may span from simple tasks as discovering dates, names and locations, to the really complex AI task of associative and analogical reasoning.

Finally, the disclosure of the personal information and identity is a manifestation of personality: open, extrovert people (and VHs) may interweave more their story with personal references than introvert ones.

An interesting question is that a VH's 'personal history' may be also adapted to a situation, or a given user (group), not only its conversational style as suggested for robots [Dautenhahn, 2004] and VHs [Ruttkay et al. to appear]. However, consistency within different interaction sessions with the same user (group) should be taken care of.

4.3 Conversational style

There is much to be exploited as of the conversational style of VHs. To begin with, variety should be aimed at in language usage. The variations should be modulated according to the identity of the VH and to the characteristics of the user. For instance, the communicative act 'greeting' should be realized differently by the VT, depending on the age of the user, weather addressing him the first time or already acquainted. Within these constraints still remains space for a couple of different utterances to choose from, including special, individual language usage.

The conversational style should also reflect the role and personality of the VH. For instance, if a VT is to be in an assistant role rather than a tutor, more informal language usage is appropriate, than in case of a physiotherapist consultant.

Similar to their 'perfect and spotless' embodiment, today's VHs talk a 'perfect and spotless' language. This is contrary to real-life conversation, where people abandon or correct sentences, use pauses and non-speech elements interwoven with words and sentences, as we have experienced, as we have also noticed in studies of conversation in meetings and in talk shows [Heylen and Op den Akker, 2006]. These 'imperfections' are not to be seen as weaknesses of real-life speech to be eliminated from the repertoire of VHs. Just the opposite, they do carry subtle information about the person (personality, level of knowledge and expertise in a filed), about the mental processes (e.g. thinking, recalling information) and conversational state (e.g. filled pauses indicate the intention to keep the floor in the conversation), or have functions related to the content, such as a filled pause used in front of a for the user negative answer to decrease the disappointment, or a pause to highlight difficult or important piece of information to come.

'Imperfect' language usage is characteristic especially in the case of the VT, where we are currently experimenting with speech elongation and alignment strategies which are beyond the normal customs, but are convenient in explaining rhythmic motions. One problem we have bumped into that TTS engines do not provide access to such 'beyond normal' control of timing.

4.4 Humor and laughter

In [Ekman, 2001] the various functions of smiles are discussed. It is generally agreed that natural interaction between humans and virtual humans requires models from which these functions can emerge. For example, in a study on visual cues for feedback the smile turned out to be the strongest cue for affirmative feedback [Granström et al., 2002]. Smiles are important in regulating interactions, but humor is also an important factor. In some environments we discussed (virtual dancer, virtual trainer and virtual conductor) it is quite natural to expect nonverbal humor in the interaction between virtual human and his human partners. In particular when one of the partners in the interaction fails to do things 'right' he or she can recover by doing something funny and unexpected or, the other way around, his or her partner can give someone the opportunity to recover by a humorous nonverbal reaction.

The role of verbal humor, with the aim to design virtual humans that are able to generate and understand verbal humor, in conversations, task-related interactions, meetings, and education, is discussed in [Nijholt, 2007]. Generating and understanding verbal humor and using humor in an appropriate way during an interaction requires natural language and common sense understanding by computers that is too far away from current research achievements in artificial intelligence. However, especially in situations as mentioned above, the combination of limited verbal intelligence, knowledge about the task or the goal of the interaction, and

the possibility to use nonverbal means can be employed by a virtual human to generate humorous acts at appropriate moments during an interaction. Adding laughter during such interactions is another issue that needs to be addressed [Trouvain and Schröder, 2004].

4.5 Here and today - situatedness

VHs hardly give the impression that they know about the time and situation they converse in with their user. Some VHs do reflect the time of the day by choosing an appropriate greeting. But much more could be done: keeping track of the day, including holidays, and commenting accordingly, providing 'geographical update' capability when placing a VH-enabled service in a location, endowed may be some social and political information about the place. Imagine a VT who knows that it is today a public holiday in Italy where the given VT is 'active'. Some special words to the user keeping up her exercise scheme on a holiday would be appropriate. But on a tropical summer day, the heat may lead the VT to revise its strategy, remind the user the necessity of drinking, or even shorten the exercises, or suggest doing it in the morning.

The identity of the user may be a source of further situatedness. As a minimum, a VH should 'remember' earlier encounters with the user. Asking the name or telling the same piece of small talk to the same person each time is disappointing. But how nice it sounds if a VT refers to yesterday's performance, knows of the user's religion does not allowing her to do exercises on Saturday, greets her specially on her birthday.
Finally, in order to perceive a VH as 'present', the VH must have means to gather information about the user and react to it. To begin with, the mere presence of the user and her identity should be detected, and her task-related performance should be monitored. But think of a real trainer or tutor, who would very likely comment on changes like not wearing glasses, change in hair style, being sunburn or showing signs of a cold. A Virtual Trainer could do similar comments.

5 Discussion

We have argued that there are dimensions to be still exploited to turn VHs more life-like, entertaining and in cases effective. We discussed the potentials of:

- making VHs look more individual and imperfect, may be configured to a given user's preferences;
- endowing VHs with identity and personal history;
- grounding VHs to the geographical and sociological place and time of the application being used;
- taking care of styled and natural conversation with phenomena of 'imperfections' reminiscent in real life.

The above features do not require, first of all, further perfection of single or multi-modal communication, but they do pose challenges on modeling mental capabilities like associative storytelling or require further socio-psychological studies of the nature and effect of social conversation in task-related situations. What is needed, for several of the above enrichments, is multi-signal perception, first of all, vision, of the conversant of a VH.
This, however, leads to a more general issue: the relationship of VHs to real ones. In our discussion we recalled examples from real human practice, which would be beneficial to endow the communicative capabilities of VHs. Due to the novelty of the field and the many parameters influencing the judgment of a VH, we cannot make conclusions as of the following questions:

- Which (as many as possible, ideally all?) phenomena of real human behaviors should be reproduced by VHs?
- How to exploit the 'beyond human' possibilities of VHs, both in perception and mental capabilities?

For both issues, dedicated evaluation studies are needed to put together a huge jigsaw image. It is clear already that the objective to engage the user in an activity and to perform a task well and efficiently may require different VH design, along several dimensions. Also, the application context (real-fictional) puts the user in different frame of mind to judge the VH, On the other hand, even less is known of the judgments of non-human capabilities of VHs. For example, it has turned out that a VH could 'read from the eye' of the user better than most of the people are capable of. What to do with such a super-human power of a VH? Or, another example is the reasoning speed and capability of a VH: do people take it as natural (from a VH) that he can recall multiple telephone books? Or should he 'fake' the human limitation of recalling data in a register?
How to get away with shallow, or not deep/complete enough, models?
Finally, we mention arts as an exploited source of design principles for VHs. Arts, especially (portrait) painting, animation and theater can provide elicit knowledge about human communication, facial and gesture expressions, reflection of personality and emotions in nonverbal signals and speech. Further on, some principles of realizing certain effects, may be by non-realistic features, can be beneficial for enhanced expressivity. The cartoonish exaggerated effects in gesturing can underline e.g. personality characteristics present in real-life speech. One step further, one envisions the next-generation, ideal VHs as seamlessly integrating elements of practices from human conversational behavior, the enhanced interaction and reasoning capabilities of computer technology and the expressivity and aesthetics of arts.

References

[Abaci et al., 2004] T. Abaci, R. de Bondeli, J. Ciger, M. Clavien, F. Erol, M. Gutierrez, S. Noverraz, O. Renault, F. Vexo and D. Thalmann. Magic Wand and Enigma of the Sphinx. *Computers & Graphics*, Volume 28, Issue 4, August 2004, 477-484.

[AgentLand, 2006] AgentLand, http://www.agentland.com/

[Badler, 2002] N. Badler. LiveActor: A virtual training environment with reactive embodied agents. Workshop on *Intelligent Human Augmentation and Virtual Environ-*

ments, University of North Carolina at Chapel Hill, Oct. 2002.

[Bailenson and Yee, 2005] J. N. Bailenson and N. Yee. Digital Chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science*, 16, 814-819. 2005.

[Baylor and Ebbers, 2003] A. L. S. Baylor and S. Ebbers. The Pedagogical Agent Split-Persona Effect: When Two Agents are Better than One. Paper presented at the *ED-MEDIA*, Honolulu, Hawaii, 2003.

[Bernsen et al., 2004] N.O. Bernsen, M. Charfuela` n, A. Corradini, L. Dybkjær, T., Hansen, S. Kiilerich, M. Kolodnytsky, D. Kupkin and M. Mehta. First prototype of conversational H.C. Andersen. International Working Conference on *Advanced Visual Interfaces (AVI'2004)*, Gallipoli, Italy, May 2004, ACM, New York, 2004, 458–461.

[Bickmore and Cassell, 2000] T. Bickmore and J. Cassell: Small Talk and Conversational Storytelling in Embodied Interface Agents. In: Proceedings of the *AAAI Fall Symposium on Narrative Intelligence*, 87–92, 2000.

[Bos, 2006] P. Bos, D. Reidsma, Z.M. Ruttkay, and A. Nijholt. Interacting with a Virtual Conductor. In: 5th *International Conference on Entertainment Computing*, LNCS 4161, Springer Verlag, Cambridge, UK, 25-54, 2006.

[Cassell et al., 2000] J. Cassell, J. Sullivan, S. Prevost, and E. Churchill (Eds.), *Embodied Conversational Agents*, MIT Press, 2000.

[Chi et al., 2000] D.M. Chi, M. Costa, L. Zhao and N.I. Badler. The EMOTE Model for Effort and Shape. *Siggraph 2000, Computer Graphics Proceedings*, ACM Press / ACM SIGGRAPH / Addison Wesley Longman, 173-182, 2000.

[Dautenhahn, 2004] K. Dautenhahn. Socially Intelligent Agents in Human Primate Culture. In: *Agent Culture: Human-Agent Interaction in a Multicultural World*, 45-71 (Chapter 3), R. Trappl and S. Payr (Eds.), Lawrence Erlbaum Associates, 2004.

[DiPaola 1991] S. DiPaola. Extending the Range of Facial Types. *Journal of Visualization and Computer Animation*, Vol.2, No. 4, 129-131. 1991.

[Ekman 1989] P. Ekman. The argument and evidence about universals in facial expressions of emotion. In: Wagner, H., Monstead, A. (Eds.): *Handbook of Social Psychology*. John Wiley, Chic ester, 1989, 143-146.

[Ekman, 2001]. P. Ekman. Telling Lies. Clues to deceit in the marketplace, politics, and marriage. New York and London: W.W. Norton and Company 2001 (reissued with a new chapter).

[Egges et al. 2004] A. Egges, T. Molet and N. Magnenat-Thalmann. Personalized Real-Time Idle Motion Synthesis. *Computer Graphics and Applications*, 12th Pacific Conference on (PG'04), 121-130, 2004.

[Granström et al., 2002]. B. Granström, D. House, and M.G. Swerts. Multimodal Feedback Cues in Human-Machine Interactions. In B. Bel and I. Marlien (Eds.), Proceedings of the *Speech Prosody 2002 Conference*, Aixen-Provence: Laboratoire Parole et Langage, 347–350, 2002.

[Gratch and Marsella, 2001] J. Gratch and S. Marsella. Tears and fears: Modeling emotions and emotional behaviors in synthetic agents. In Proceedings of the *Fifth International Conference on Autonomous Agents*, 2001.

[Gratch et al., 2002] J. Gratch, J. Rickel, E. Andre, N. Badler, J. Cassell, and E. Petajan. Creating interactive virtual humans: Some assembly required. *IEEE Intelligent Systems*, 2002.

[Gustafson et al., 2004] J. Gustafson, L. Bell, J. Boye, A. Lindström, and M. Wiren. The NICE Fairy-tale Game System. Proceedings of *SIGdial 04*, Boston, 2004.

[Hartmann et al., 2005] B. Hartmann, M. Mancini, S. Buisine, and C. Pelachaud. Design and Evaluation of Expressive Gesture Synthesis for Embodied Conversational Agents. *AAMAS'05*, Utrecht, July 2005.

[Hayes-Roth and Doyle, 1998] B. Hayes-Roth and P. Doyle. Animate characters. *Autonomous Agents and Multi-Agent Systems*, Volume 1, Number 2, 1998.

[Heylen and Op den Akker, 2006] D. Heylen and R. op den Akker. Investigations into the distribution of back-channels in argumentative multi-party discourse and their functional determinants. Manuscript, HMI Research group, University of Twente, September 2006.

[Isbister et al., 2000] K. Isbister, H. Nakanishi, T. Ishida and C. Nass. Helper agent: Designing an assistant for human-human interaction in a virtual meeting space. Proceeding of CHI'2000, 57-64.

[Isbister and Doyle, 2004] K. Isbister and P. Doyle. The Blind Man and the Elephant Revisited: A Multidisciplinary Approach to Evaluating Conversational Agents. In (Z. Ruttkay and C. Pelachaud, Eds.) *From Brows to Trust: Evaluating Embodied Conversational Agents*, Volume 7, Human Computer Interaction series, Kluwer Press. 2004.

[Koda and Maes, 1996] T. Koda and P. Maes. Agents with faces: The effects of personification of agents. Proc. of HCI'96, 1996.

[Liang et al., 2002] L. Liang, H. Chen, Y. Q. Xu and H. Y. Shum, Example-Based Caricature Generation with Exaggeration, *in Proc. 10th Pacific Conf. on Computer Graphics and Applications*, 2002.

[Kopp et al., 2003] S. Kopp, B. Jung, N. Lessmann, and I. Wachsmuth. Max--A Multimodal Assistant in Virtual Reality Construction. *KI-Künstliche Intelligenz* 4/03: 11-17, 2003.

[Mateas and Stern, 2003] M. Mateas and A. Stern. Facade: An Experiment in Building a Fully-Realized Interactive

Drama. *Game Developer's Conference:* Game Design Track, San Jose, California.

[Mori 1982] M. Mori. *The Buddha in the Robot.* Charles E. Tuttle Co., 1982.

[Nijholt, 2007] A. Nijholt. Conversational Agents and the Construction of Humorous Acts. Chapter 2 in: *Engineering Approaches to Conversational Informatics*. T. Nishida (Ed.), John Wiley & Sons, Chichester, England, 2007, to appear.

[Noot and Ruttkay, 2004] H. Noot, Zs. Ruttkay: Style in Gesture. In: A. Camurri, G. Volpe (Eds.), *Gesture-Based Communication in Human-Computer Interaction*. LNCS 2915, Springer-Verlag, 2004.

[Payr and Trappl, 2004] S. Payr, R. Trappl (Eds.). *Agent Culture - Human-agent Interaction in a Multicultural World*. Lawrence Erlbaum Associates, 2004.

[Plantec, 2004] P. Plantec. Virtual Humans. *AMACOM*, 2004.

[Poggi et al., 2001] I. Poggi, C. Pelachaud, and B. De Carolis. To display or not to display? Towards the architecture of a reflexive agent. In *Proceedings of the 2nd Workshop on Attitude, Personality and Emotions in User-adapted Interaction. User Modeling 2001*, 13-17. 2001.

[Prendinger and Ishizuka, 2001] H. Prendinger and M. Ishizuka. S*ocial role awareness in animated agents*, Proc. of Autonomous Agents Conference, 270-277, Montreal, Canada. 2001.

[Prendinger and Ishizuka, 2004] H. Prendinger and M. Ishizuka (Eds.). *Life-Like Characters. Tools, Affective Functions, and Applications*, Cognitive Technologies Series, Springer, Berlin Heidelberg, 2004.

[Reidsma et al., 2006] D. Reidsma, H. Van Welbergen, R.W. Poppe, P. Bos, and A. Nijholt. Towards Bidirectional Dancing Interaction. Proc. of 5th *International Conference on Entertainment Computing*, LNCS 4161, Springer-Verlag, 2006, 1-12, 2006.

[Rehm and Andre, 2005] M. Rehm and E. Andre. Catch me if you can -- exploring lying agents in social settings. Proc. of the Int. Conf. on Autonomous Agents and Multiagent Systems, 937-944, 2005.

[Ruttkay and Pelachaud, 2004] Z. Ruttkay and C. Pelachaud (Eds.). *From Brows to Trust. Evaluating Embodied Conversational Agents*. Kluwer's Human-Computer Interaction Series – volume 7, 2004.

[Ruttkay et al., 2006a] Z. M. Ruttkay, D. Reidsma, and A. Nijholt. Human Computing, Virtual Humans and Artificial Imperfections. *ACM-ICMI'06*, Banff, Canada, 2006.

[Ruttkay et al., 2006b] Zs. Ruttkay, J. Zwiers, H. van Welbergen, and D. Reidsma. Towards a Reactive Virtual Trainer. Proc. of *IVA 2006*, LNAI 4133, Springer-Verlag, 292-303. 2006.

[Ruttkay et al., to appear] Zs. Ruttkay, C. Pelachaud, I., Poggi, and H. Noot. Exercises of Style for Virtual Humans. In: L. Canamero, R. Aylett (Eds.), *Advances in Consciousness Research Series*, John Benjamins Publishing Company, to appear.

[Stronks et al., 2002] 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.

[Thorisson, 1996] K. Thorisson. *Communicative Humanoids: A Computational Model of Psychosocial Skills*. PhD thesis, MIT Media Laboratory, 1996.

[Trouvain and Schröder, 2004] J. Trouvain and M. Schröder. How (Not) to Add Laughter to Synthetic Speech. In: Proc. Workshop on *Affective Dialogue Systems (ADS 04)*, E. André (Ed.), LNAI 3068, Springer Verlag, Kloster Irsee, Germany, 2004, 229-232.

[Van Moppes 2002] V. Van Moppes. *Improving the quality of synthesized speech through mark-up of input text with emotions*. Master Thesis, VU, Amsterdam, 2002.

[Wachsmuth and Kopp, 2002] I. Wachsmuth and S. Kopp. Lifelike Gesture Synthesis and Timing for Conversational Agents. In Wachsmuth & Sowa (eds.), *Gesture and Sign Language in Human-Computer Interaction*, International Gesture Workshop (GW 2001), 120-133, LNAI 2298, Springer-Verlag, 2002.