Engaging in a Conversation with Synthetic Characters Along the Virtuality Continuum

Elisabeth André, Klaus Dorfmüller-Ulhaas, and Matthias Rehm

Multimedia Concepts and Applications, University of Augsburg, Germany
andre@informatik.uni-augsburg.de

Abstract. During the last decade research groups as well as a number of commercial software developers have started to deploy embodied conversational characters in the user interface especially in those application areas where a close emulation of multimodal human-human communication is needed. Most of these characters have one thing in common: In order to enter the user's physical world, they need to be physical themselves. The paper focuses on challenges that arise when embedding synthetic conversational agents in the user's physical world. We will start from work on synthetic agents that populate virtual worlds and anthropomorphic robots that inhabit physical worlds and discuss how the two areas need to be combined in order to populate physical worlds with synthetic characters. Finally, we will report on so-called traversable interfaces that allow agents to cross the border from the physical space to the virtual space and vice versa.

1 Introduction

The objective to develop more human-centered, personalized and at the same time more entertaining interfaces immediately leads to the metaphor of an embodied conversational agent that employs gestures, mimics and speech to communicate with the human user. Incarnations of such characters differ widely in type and amount of embodiment – starting from simplistic cartoon-style 2D representations of faces, fully embodied virtual humans in 3D virtual worlds to physically embodied androids co-habiting the user's real world (see [1] for an overview). Despite of their variety, most of these characters have one thing in common: In order to get in touch with a user in the physical space, the characters have to be physical themselves.

Following [2], we may classify the contact between synthetic and human agents according to a "virtuality continuum" (see Fig. 1). At one extreme, we find android agents that are completely integrated in the user's physical world and even allow for physical contact with the user. Mel, a robotic penguin developed by Sidner and colleagues [3] (see image 1 in Fig. 1), is one of the most sophisticated physical agents that engages in face-to-face communication with a human user. At the other extreme, there are purely virtual environments that are populated by human and synthetic agents. A prominent example is the pedagogical agent Steve [4] (see Image 4 in Fig. 1). Steve is aware of the user's presence in the virtual space, monitors her actions and responds to them, but has no access to the external world. That is it is only able to perceive user actions that are performed in the virtual space. In between, we find a new generation of characters that inhabit a world in which virtual and digital objects are smoothly integrated. In these applications, projections of virtual characters overlay the user's physical environment or projections of real persons are inserted into a virtual world. For instance, Cavazza and colleagues [5] propose a magic mirror paradigm which puts the user both in the role of an actor and a spectator by inserting the user's video image in a virtual world that is populated by synthetic agents (see Image 3 in Fig. 1). Kruppa and colleagues [6] focus on the reciprocal problem, namely how to populate the user's physical environment with synthetic agents. Their character is capable of freely moving along the walls of a real room by making use of a steerable projector and a spatial audio system. The Virtual Augsburg project aims at an even tighter integration of a character into a physical world (see [7]). In this application, a synthetic character called Ritchie jointly explores with the user a table-top application that combines virtual buildings of the city center of Augsburg with a real city map being laid out on a real table. Augmented Reality applications are characterized by a tight spatial connection between physical and virtual objects. This includes the correct projection of occlusions as well as the generation of realistic shadows falling onto real objects. The second image in Fig. 1 shows the character Ritchie when entering the Multimedia Lab. As shown in the image, the real door frame partially covers, for instance, the virtual character which in turn occludes the wall in the background.

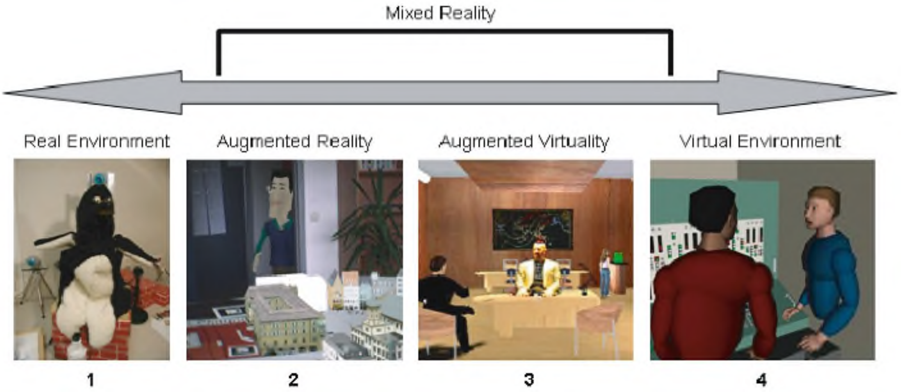

Fig. 1. Milgram's Diagram of Virtuality Continuum Adapted to Embodied Conversational Characters: the robotic penguin developed by Sidner and colleagues [3] at MERL (image 1), the virtual character Ritchie entering the Multimedia Interfaces Lab at Augsburg University (image 2), Marc Cavazza acting as "Goldfinger" in an Augmented Virtuality application developed by his team at Teaside University [5] (image 3), the pedagogical agent Steve developed by Rickel and Johnson at ISI [4] (image 4)

Most work so far has concentrated on the design and implementation of conversational agents at the two extremes of the Virtuality Continuum. The objective of this paper is to investigate which challenges arise from embedding synthetic conversational characters in the user's physical world. We will start from work on synthetic agents populating virtual worlds as well as anthropomorphic robots inhabiting physical worlds and discuss how the two areas need to be combined in order to enable characters to enter worlds between the two extremes of the Virtuality Continuum and engage in a conversation with a human user. In particular, we are interested in grounding, i.e. how to establish a common understanding between a user and a synthetic agent of what is being said and meant at the various levels of the Virtuality Continuum. Furthermore, we will describe challenges resulting from so-called traversable interfaces which allow a character to cross the border from the real to the virtual world and vice versa.

2 Perceptual Attention

Face-to-face communication does not take place in an empty space, but should be linked to the surroundings of the conversational partners. An embodied agent may acquire knowledge about its surroundings by a sensory mechanism and/or by consulting a model of the interaction space that might be mentioned during the interaction. In general we can distinguish between synthetic vision and real vision approaches. The first refers to models of attention and perception in the virtual world whereas the latter comprises efforts to recognize and interpret nonverbal input from the real world, such as objects in the interaction space or a user's gestures and head movements.

Synthetic vision approaches model (part of) the human perceptual process in an agent which is situated in a virtual environment. The agent is supplied with (virtual) visual sensors which are used to extract visual information about the environment. A characteristic feature of all approaches in this area is the idea to enable only limited access to information about the world. On the one hand, this allows for a more natural attention behavior of the agent because it is no longer omniscient but has to actively look around to gather necessary information. On the other hand, it is important to reduce the processing load of the agent. If the world becomes very large, the price to pay for omniscience is the need for much computational power. Another feature is the distinction between a bottom-up and a top-down control of attention. Bottom-up control is involuntary and triggered e.g. by the saliency of an object (bright color, movement) whereas top-down control is a goal-directed volitional act, e.g. finding a grey ball which will focus attention on round objects with less bright colors.

Peters and O'Sullivan [8] describe a model for bottom-up visual attention that attempts to mimic the low-level, automatic mechanisms of early visual processing which rely on salience features and draw attention to salient locations in the input. Thus, their agent is capable of idle or spontaneous looking which takes place when there is no task to perform. It can also be employed in case of a necessary task interruption, e.g. in case of a threat, such as a moving car. Their model combines database queries (on the velocity of objects) with a visual sensor. Due to the restricted access to the scene database, the model is enhanced with an object-based short-term memory. Input to the memory component are simplified images without textures and lighting that simulate peripheral and foveal vision. In the case of peripheral vision only groups of objects are recognized.

Hill [9] proposes a two step model of perceptual attention for virtual agents that focuses mainly on the problem of object grouping to prevent a conceptual overload. By grouping similar objects, they can be treated as one object. He distinguishes between an automatic process of grouping which he describes as pre-attentive and a process of attention control. Automatic grouping segments the input and groups objects according to Gestalt principles, such as proximity or similarity. Groups are created, joined or split making use of the center of mass defined by similar objects and a fixed grouping radius that originates in this center of mass. Attention control then is a voluntary process that can be triggered in two ways: (i) bottom-up by salient external stimuli, such as luminance, color or motion, (ii) top-down by the cognitive system in which case attention is task-oriented, e.g., for searching or tracking objects.

Chopra and Badler [10] present a psychologically-motivated framework for automating a character's visual attention. It implements a model of eye behavior where repetitive cycles of eye movements constitute pre-defined patterns of looking behaviors that are associated with motor activities (walk, reach, etc.) as well as with cognitive actions (search, track, etc.). The authors emphasize the importance of the agent's task for its attentive behaviors. Without a task, the agent exercises spontaneous looking where attention is easily captured by peripheral events. In case the agent engages in a task, attention is more focused rejecting such an exogenous capturing of attention.

Whereas synthetic vision approaches try to model human perceptual and attentive processes in the virtual environment, computer vision approaches are rather concerned with capturing actions and events in the real world. In contrast to synthetic vision, the real world is much less predictable, more complex and dynamic as the virtual world. Thus, the goal cannot be a simulation of the human perceptual process, but the recognition of certain real world features to further a more natural interaction. Examples of such recognition tasks are head tracking which allows the agent to maintain mutual gaze or interpret the user's looking behaviors, gesture recognition and interpretation for identifying points of interest the user indicates, or recognition of objects that can be used as shared references.

The virtual agent Max (e.g. see [11] or [12]) employs a number of different sensors to collect information from the user aiming at the detection of context-sensitive foci. The face recognition system relies on detecting skin colored objects in the upper regions of a camera image allowing the agent to focus on these salient regions and giving the user the impression that the agent is maintaining mutual gaze with the user. Capturing the user's gestures just from a visual camera image is error-prone, resulting in the use of data gloves and a corresponding tracking system.

Agents that inhabit Augmented Realities need to be aware of the physical as well as the digital space. As a consequence, we have to fuse the results from the synthetic and real vision processes. One specific problem is the different information density for the real and for the virtual world. Perception and attention of the virtual part of the scenario can be modelled to whatever degree is desirable because all the information about the virtual world is in principle available. Perceiving and attending to the real world faces severe technical problems and is thus always limited and deficient unless the surrounding real world is static and can thus be modelled as well. Furthermore, new methods are required for calculating attentional prominence in an Augmented Reality which take into account both physical and digital object features.

3 Engagement Cues in Conversation

It does not make much sense to implement mechanisms for perceptive behaviors if the agent is not able to show its awareness for the environment. In this paper, we focus on attention in a conversation as an important means to show engagement in a dialogue. Sidner and colleagues [3] refer to engagement as “the process by which two (or more) participants establish, maintain and end their perceived connection during interactions they jointly undertake”. Engagement behaviors may be directed to the conversational partner as well as the conversational environment. Clark [13] consider attention to the speaker as an important signal of understanding by the addressee. Furthermore, Nakano and colleagues [14] observed that shared attention to an object may be interpreted as a sign of common understanding for the task at hand.

3.1 Engagement Cues to the Conversational Environment

To indicate attention to objects in the environment, Clark [15] proposes several directing-to and placing-for behaviors.

Directing-to behaviors may be realized by a number of verbal and non-verbal means including demonstrative pronouns, eye gaze or pointing gestures. In an Augmented Reality, directing-to behaviors may relate to both virtual and physical objects. For instance, the character Ritchie in Image 2 in Fig. 1 might generate a mixed deictic utterance, such as “the building in the middle of the table” to help the user to localize a digital object in the physical space. Existing algorithms for the analysis and generation of deictic referring expressions usually start from the assumption that all reference objects are located in the same space - either virtual (see [16]) or physical (e.g. see [17]). To extend these algorithms to the broader context of mixed realities, real and digital objects have to be localized within a uniform coordinate system. Furthermore, we need to take into account the real and digital properties of objects as well the position of real and virtual conversational partners. For instance, to generate a deictic utterance, such as “the building on the table in front of you” in Image 2 in Fig. 1, the system needs to know the relative position of a physical object (the table), a digital object (the building) and the user in the real space.

Placing-for behaviors differ from directing-to behaviors by moving objects in the addressee’s focus of attention. For instance, a customer may place a product on the counter to signal the cashier that she wishes to buy the product. Interestingly, such behaviors serve to provide common ground between the customer and the cashier regarding the action to be performed. That is they are usually understood without any additional verbal comment, just by considering the current context. Augmented Reality interfaces offer novel forms of placing-for behaviors that are situated in a physical space. For instance, a synthetic character may place a virtual object on a real table to direct the user’s attention to this object. Vice versa, a human user may position physical objects in the character’s focus of attention. In the Virtual Augsburg application, the user may position a cardboard box in the physical scene to signal the character where it should move (see Fig. 2). A frustum of pyramid with markers attached to each side is used to facilitate the recognition of the position and orientation of the cardbox employing the user’s head-worn camera.

Fig. 2. Placing a Physical Object in the Real Space to Attract a Character's Attention

3.2 Engagement Cues to the Conversational Partner

A number of researchers focus on eye gaze as one of the most important means to indicate engagement in a dialogue. According to Kendon [18], we can distinguish between at least four functions of seeking or avoiding to look at the partner in dyadic interactions: (i) to provide visual feedback, (ii) to regulate the flow of conversation, (iii) to communicate emotions and relationships, (iv) to improve concentration by restriction of visual input. Concerning the listener, Argyle and Cook [19] show that people look nearly twice as much while listening than while speaking. Compared to dyadic conversations, we know little about gaze behavior in multiparty interaction. Vertegaal and colleagues [20] describe a study of the looking behavior in a four-party interaction. Subjects looked about 7 times more at the individual they listened to than at others. They looked about three times more at the individual they spoke to than at others. In accordance with Sidner et al. [3] or Nakano et al. [14], they conclude that gaze, or looking at faces, is an excellent predictor of conversational attention in multiparty conversations.

Empirical studies of engagement in human-human conversation have inspired the implementation of engagement behaviors for embodied conversational agents. Nakano and colleagues [14] developed a engagement model for the kiosk agent Mack that provides route directions for a paper map. The agent uses gaze as a deictic device as well as a feedback and turn taking mechanism. Based on an analysis of human-human conversation, Sidner and colleagues [3] implemented a conversational robot that is able to track the face of the conversational partner and adjusts its gaze towards him or her. Even though the set of communicative gestures of the robot was strongly limited, an empirical study revealed that users indeed seem to be sensitive to a robot's conversational gestures and establish mutual gaze with it. Nakano and Nishida [21] present a synthetic agent which recognizes the user's gaze direction as a cue to non-verbal engagement. The agent is situated on a large screen in front of a static background image which offers a number of points of interest that can be brought up during the conversation. The conversation is influenced by the conversational and the perceptual situation, i.e., the user's focus of attention given by his/her gaze direction.

In order to enable an agent to reason about another agent's attentive behaviors, Peters [22] proposes to rely on a theory of mind. The model by Baron-Cohen [23] seems to offer a promising approach since it relates gaze perception to higher level cognitive processes. Based on this model, Peters proposes the implementation of a synthetic vision module that considers to what extent the orientation of an agent's eyes, head and body are directed towards the conversational partner in order to determine its attention towards the conversational partner.

In the Virtual Augsburg Scenario, the user's head position and orientation is tracked to monitor his or her focus of attention. The tracking is enabled by a small camera fixed on the user's head-mounted display (see Fig. 3) and a marker tracking software running on a standard PC.[1] In this way, the character may intervene if the user is loosing interest. Nevertheless, the character is only able to reason about the user's attention towards the digital world and a strongly limited section of the real world (essentially the table) so far.

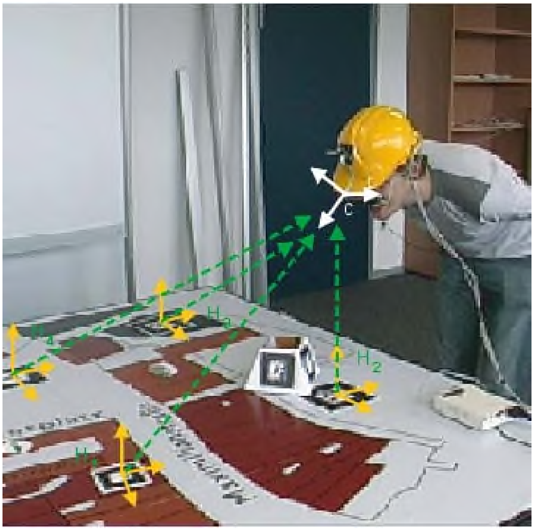

Fig. 3. Tracking System for Virtual Augsburg

4 Conversational Locomotion

While most work on face-to-face communication focuses on the generation of gestures, mimics and speech, hardly any research has addressed the problem of conversational locomotion. Nevertheless, face-to-face communication often requires the participants to change their position in space. In a party-like scenario, people dynamically join or

---

[1] We have started with the wide-spread AR Toolkit. However, we have created our own tracking system which relies on improved marker detection and pose estimation processes.

leave groups with which they engage in a communication. Furthermore, people often move to another location in order to be closer to objects they wish to talk about.

An early approach to coordinate locomotive , gestural, and speech behaviors for an embodied conversational agent has been presented by Lester and colleagues [24]. Peters [22] discusses the use of locomotive behaviors, directed gestures, body orientation and gaze to express different levels of attention to objects and persons. Thalmann and colleagues [25] concentrate on the simulation of social navigation behaviors in virtual 3D environments including the social avoidance of collisions, intelligent approach behaviors, and the calculation of suitable interaction distances and angles. The work is based on an operationalization of empirically-grounded theories of human group dynamics, such as Kendon's group formation system [18]. To enable such a behavior in an Augmented Reality, the agent needs to navigate through a physical space, thereby avoiding collisions with physical and digital objects.

The question arises, however, of whether humans follow similar social group dynamics when interacting with synthetic agents in Virtual or Augmented Realities. To explore social navigation behaviors of humans in immersive 3D worlds, Bailenson and colleagues [26] have conducted a number of empirical studies which revealed that human subjects apply similar behavior patterns in a virtual world as in a physical world. For instance, the size and the shape of the personal space around the virtual humans resembled the shape of the space that people leave around real, nonintimate humans. Another interesting result was the observation that the social navigation behavior depends on whether the agent is controlled by a computer or by a real human. They found out that people gave an avatar more personal space than an autonomous agent. In a second experiment, they examined what happens when virtual humans walk and violate the personal space of the participants. The experiment showed that human subjects avoided the virtual human more when it was obviously computer-controlled than an avatar that was controlled by another human.

In the experiments above, the user was completely immersed in the real world. Naturally, the question arises of whether the same findings hold for social navigation behaviors between human and virtual agents in an Augmented Reality. Most likely the navigation behavior will depend on how smoothly the character is integrated in the physical world. In particular, deficiencies of the employed display technology might disturb the user's sense of sharing a space with the character and lead to unexpected or little natural behaviors.

5 Traversable Interfaces

The preceding sections started from scenarios in which characters were bound to a specific space: a real, virtual or augmented environment. An intriguing challenge is the realization of so-called traversable interfaces [27] that allow human and synthetic agents to cross the border from the digital world to the real world and vice versa. One advantage of traversable interfaces lies in the fact that it allows users to move to another interaction space if the information presentation requires it. An example of a traversable interface is shown in Fig. 4. The character Ritchie is first exploring with the user a virtual 3D representation of the city of Augsburg. It then moves to the real space to study a miniaturized model of the city of Augsburg laid out on a table.

Fig. 4. Moving from the Real to the Virtual World an Vice Versa

In research on virtual reality, a higher sense of presence has been considered as a critical success for many applications, such as virtual meeting rooms or training applications. Witmer and Singer [28] define *presence* as "the subjective experience of being in one place or environment, even when one is physically situated in another". Usually, it will take the users some time to get a sense of presence for a new world. Traversable interfaces need to provide for means that enable the user a smooth transfer to another world. It may help the user if the system announces the transfer to another interaction space, for example, by employing beam techniques as in Star Trek (see [29]). In order to help the users to orient themselves, we make use of a virtual companion as a uniform interaction metaphor that accompanies them not only in the digital, but also in the real space. For instance, on the left-hand side of Fig. 4, user and character explore a virtual 3D model of Augsburg from a pedestrian's angle. In our case, there is no visual representative of the user in the virtual 3D space which is supposed to follow the agent looking over its shoulder.[2] On the right-hand side of Fig. 4, user and character are both situated in the physical world and study the city model of Augsburg from a bird's eye view. In both cases, the user perceives the scene in a similar way as the character does in order to make it easier for her to share experiences with it.

In order to establish a temporally persistent relationship with the user, the agent should maintain a memory of joint experiences in the different worlds. For instance, when visiting 3D model of the town hall, the agent should remember what it told the user about the corresponding table-top model. To allow for references from one world to the other, we need to maintain a dialogue history across interaction spaces. To our knowledge, no research so far has investigated how mechanisms for the generation and analysis of anaphora need to be adapted to cope with traversable interfaces. So-called alignment problems may arise if the agent produces referring expressions for objects in

---

[2] We still need to investigate whether the missing embodiment of the user in the 3D space leads to a too drastical lost of presence. In this case, it might help to visualize parts of the user's body, such as her forearms.

the new interaction space which the user is not able to resolve because she has not yet succeeding in developing a sense of presence for the new world.

We are currently preparing an empirical study to investigate in how far the use of a virtual companion may help the user to move from one space to the other. In particular, we are interested in the question of whether the use of a virtual character contributes to a more coherent perception of the interaction.

6 Conclusion

In this paper, we have presented a new generation of synthetic characters that are no longer bound to a flat screen, but able to enter a physical world and to engage in a conversation with a human user. Users and characters do not inhabit separated spaces, but share an informational and physical reality that is augmented by digital objects. As a consequence, communication has to take into account both the physical and the digital context. New forms of deixis are enabled by the manipulation of objects and movements of characters in the physical space. Further challenges arise from the realization of so-called traversable interfaces that allow human and synthetic agents to cross the border from the digital world to the real world and vice versa. Obviously, mixed realities put high demands to the agents' perceptive skills requiring to integrate mechanisms for real and synthetic vision. It remains a great challenge to generate plausible engagement behaviors despite of seriously impoverished perception channels.

Acknowledgement

This paper was partially funded by the EU Network of Excellence Humaine. We would like to thank Peter Rist for the graphical design of the character Ritchie. We are also grateful to Lewis Johnson, Fred Charles and Candy Sidner for giving us the permission to use their figures. The copyright for the figures is to the single institutions.

References

1. Gratch, J., Rickel, J., André, E., Badler, N., Cassell, J., Petajan, E.: Creating interactive virtual humans: Some assembly required. IEEE Intelligent Systems **17** (2002) 54–63
2. Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. IEICE Transactions on Information Systems **E77-D** (1994) 1321–1329
3. Sidner, C.L., Kidd, C.D., Lee, C., Lesh, N.: Where to look: a study of human-robot engagement. In: IUI '04: Proceedings of the 9th international conference on Intelligent user interface, New York, NY, USA, ACM Press (2004) 78–84
4. Jeff Rickel, W.L.J.: Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. Applied Artificial Intelligence **29** (1999) 343–382
5. Cavazza, M., Charles, F., Mead, S.J., Martin, O., Marichal, X., Nandi, A.: Multimodal acting in mixed reality interactive storytelling. IEEE-Multimedia **11** (2004) 30–39
6. Kruppa, M., Spassova, L., Schmitz, M.: The virtual room inhabitant. In: Proc. of the IUI-2005Workshop on Multi-User and Ubiquitous User Interfaces (MU3I), San Diego, USA (2005)

7. André, E., Dorfmüller-Ulhaas, K., Rist, T.: Embodied conversational characters: Wandering between the digital and the physical world. it - Information Technology **46** (2004) 332–340
8. Peters, C., O'Sullivan, C.: Bottom-up Visual Attention for Virtual Human Animation. In: roceedings of the Computer Animation and Social Agents (CASA) 2003, Rutgers University, New York. (2003) 111–117
9. Hill, R.: Modeling Perceptual Attention in Virtual Humans. In: Proceedings of the 8th Conference on Computer Generated Forces and Behavioral Representation, Orlando, FL. (1999)
10. Chopra-Khullar, S., Badler, N.I.: Where to look? Automating attending behaviors of virtual human characters. In: Proceedings of the third annual conference on Autonomous Agents, New York, ACM Press (1999) 16–23
11. Jung, B., Kopp, S.: Flurmax: An interactive virtual agent for entertaining visitors in a hallway. In et al., T.R., ed.: Intelligent Virtual Agents. Springer, Berlin, Heidelberg (2003) 23–26
12. Kopp, S., Jung, B., Lessmann, N., Wachsmuth, I.: Max — A Multimodal Assistant in Virtual Reality Construction. KI – Künstliche Intelligenz (2003) 11–17
13. Clark, H.: Using Language. Cambridge University Press, Cambridge (1996)
14. Nakano, Y., Reinstein, G., Stocky, T., Cassell, J.: Towards a model of face-to-face grounding. In: Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2003). (2003) 553–561
15. Clark, H.: Pointing and Placing. In Kita, S., ed.: Pointing: Where language, culture, and cognition meet. Erlbaum, Hillsdale, N.J. (2003) 1–17
16. Gorniak, P., Roy, D.: A visually grounded natural language interface for reference to spatial scenes. In: ICMI '03: Proceedings of the 5th international conference on Multimodal interfaces, New York, NY, USA, ACM Press (2003) 219–226
17. Roy, D., Hsiao, K.Y., Mavridis, N.: Mental imagery for a conversational robot. IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS **34** (2004) 1374–1383
18. Kendon, A.: Nonverbal Communication, Interaction, and Gesture. Mouton, The Hague (1981)
19. Argyle, M., Cook, M.: Gaze and Mutual Gaze. Cambridge University Press, Cambridge (1976)
20. Vertegaal, R., Slagter, R., van der Veer, G., Nijholt, A.: Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In: CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY, USA, ACM Press (2001) 301–308
21. Nakano, Y.I., Nishida, T.: Awareness of Perceived World and Conversational Engagement by Conversational Agents. In: Proceedings of the AISB 2005 Symposium on Conversational Informatics for Supporting Social Intelligence & Interaction. (2005)
22. Peters, C.: Foundations of an Agent Theory of Mind Model for Conversation Initiation in Virtual Environments. In: Proceedings of the AISB 2005 Joint Symposium on Virtual Social AGents. (2005) 163–170
23. Baron-Cohen, S.: How to build a baby that can read minds: Cognitive Mechanisms in Mind-Reading. Cahiers de Psychologie Cognitive **13** (1994) 513–552
24. Lester, J., Voerman, J., Towns, S., Callaway, C.: Deictic believability: Coordinating gesture, locomotion, and speech in lifelike pedagogical agents. Applied Artificial Intelligence **29** (1999) 383–414
25. Guye-Vuillième, A., Thalmann, D.: A high level architecture for believable social agents. Virtual Reality Journal **5** (2001) 95–106
26. Bailenson, J.N., Blasovich, J., Beall, A.C., Loomis, J.M.: Interpersonal distance in immersive virtual environments. Personality and Social Psychology Bulletin **29** (2003) 819–833

27. Koleva, B., Schnädelbach, H., Benford, S., Greenhalgh, C.: Traversable interfaces between real and virtual worlds. In: CHI ’00: Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY, USA, ACM Press (2000) 233–240
28. Witmer, B., Singer, M.: Measuring presence in virtual environments: A presence questionnaire. Presence: Teleoperators and Virtual Environments **7** (1998) 225–240
29. Rocchi, C., Stock, O., Zancanaro, M., Kruppa, M., Krüger, A.: The museum visit: generating seamless personalized presentations on multiple devices. In: IUI ’04: Proceedings of the 9th international conference on Intelligent user interface, New York, NY, USA, ACM Press (2004) 316–318