

Where Do They Look? Gaze Behaviors of Multiple Users Interacting with an Embodied Conversational Agent

Matthias Rehm and Elisabeth André

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

Abstract. In this paper, we describe an experiment we conducted to determine the user’s level of engagement in a multi-party scenario consisting of human and synthetic interlocutors. In particular, we were interested in the question of whether humans accept a synthetic agent as a genuine conversational partner that is worthy of being attended to in the same way as the human interlocutors. We concentrated on gaze behaviors as one of the most important predictors of conversational attention. Surprisingly, humans paid more attention to an agent that talked to them than to a human conversational partner. No such effect was observed in the reciprocal case, namely when humans addressed an agent as opposed to a human interlocutor.

1 Introduction

In face-to-face communication, humans employ a number of verbal and non-verbal signals to show their level of engagement in a dialogue. According to Sidner and colleagues [15] engagement “is the process by which two (or more) participants establish, maintain and end their perceived connection during interactions they jointly undertake”. The appropriate use and correct interpretation of engagement signals is a necessary prerequisite for the success of an interaction. In particular, gaze has been recognized as an important means to show engagement in a dialogue. While the listener employs gaze to indicate that s/he is paying attention to the speaker, the speaker monitors the listener’s gaze to find out whether s/he is still interested in continuing the conversation.

According to Kendon [9], 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. Kendon showed that speakers tend to look away at the beginning of an utterance and turn their attention towards the conversational partner at the end of an utterance. Regarding the listener, Argyle and Cook [1] show that people look nearly twice as much while listening (75%) than while speaking (41%).

Compared to dyadic conversations, we know little about gaze behavior in multiparty interactions. Vertegaal and colleagues [16] describe a study of the gaze behavior in a four-party interaction. Subjects looked about 7 times more at the individual they listened

to (62%) than at others (9%). They looked about three times more at the individual they spoke to (40%) than at others (12%). In accordance with Sidner et al. [15] or Nakano et al. [12], they conclude that gaze, or looking at faces, is an excellent predictor of conversational attention in multiparty conversations. Vertegaal et al. also showed that

1. People look more at the person they speak or listen to than at others.
2. Listeners in a group can still see they are being addressed. Each person still receives 1.7 times more gaze than could be expected had s/he not been addressed.
3. Speakers compensate for divided visual attention by increasing the total amount of their gazes.
4. Listeners gaze more than speakers (1.6 times).

The question arises of whether the attentive behaviors of humans change when they interact with a synthetic agent instead of another human. Most studies conducted on gaze behaviors in human-agent conversation focus on humanoid avatars representing people engaged in a conversation. This research is driven by the objective to provide the participants of video-based tele conferences with a means to establish a connection via gaze.

Colburn and colleagues [5] investigated whether natural eye gaze behaviors of an avatar that are informed by studies of human-human conversation elicit more natural eye gaze behaviors in users communicating with it. When an avatar was present, subjects spent more time looking at the screen. Even more attention was directed to the avatar when the agent relied on an eye gaze model that was informed by psychological studies on human-human conversation. Colburn and colleagues hypothesize that humans feel less shy when talking to a monitor than when talking to a real human. The effect occurred, however, only in the user-as-speaker condition which Colburn and colleagues attribute to the bad quality of the employed lip-synch mechanism.

While Colburn and colleagues concentrate on the behavioral response to avatars employing an informed eye gaze model, Garau and colleagues [7] as well as Lee and colleagues [11] investigate the effect of informed gaze models on the perceived quality of communication by means of questionnaires. Both research teams observed a superiority of informed eye gaze behaviors over randomized eye gaze behaviors. A follow-up study by Vinayagamoorthy and colleagues [17] focused on the correlation between visual realism and behavioral realism. They found that the model-based eye gaze model improved the quality of communication when a realistic avator was used. For cartoonish avatars, no such effect was observed.

Another line of research focuses on how humans respond to the gaze behaviors of autonomous conversational agents. Nakano and colleagues [12] developed a model of grounding for the kiosk agent Mack that provides route descriptions for a paper map. The agent uses gaze as a deictic device as well as a feedback and turn taking mechanism to establish a common understanding between user and agent of what is being said and meant. A preliminary study revealed that a system with a grounding mechanism seems to encourage more non-verbal feedback from the user than a system without any grounding mechanism.

Based on an analysis of human-human conversation, Sidner and colleagues [15] developed a model of engagement for a conversational robot that is able to track the user's

face and adjusts its gaze accordingly. Even though the set of communicative behaviors 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.

Summing up, it can be said that the studies found a positive effect of natural gaze behaviors on the communication between humans and synthetic agents. While the studies above focused on dyadic interactions between humans and agents, we will investigate how humans behave in a multi-party dialogue scenario with human and synthetic participants. In particular, we are interested in the following questions:

1. Do people apply different attentive behavior patterns in multi-party scenarios when *talking to an agent* as opposed to *talking to a human*?
2. Do people apply different attentive behavior patterns in multi-party scenarios when *listening to an agent* as opposed to *listening to a human*?

To investigate such behaviors, we recorded users interacting with a human and a synthetic game partner in a game of dice called Mexicali. The scenario allows us to directly compare gaze behaviors in human-human with gaze behaviors in human-agent interaction.

We suppose that humans interact with an agent in a way that roughly resembles interaction with a human. Based on [1] and [16], we assume that humans spend more time on looking at the agent when listening to it than when talking to it. Following [9], we expect similar behaviors at sentence boundaries as in human-human communication. Nevertheless, the user will probably pay more attention to the other human conversational partner since the communicative skills of the agent are strongly limited. For instance, the user might not establish frequent gaze contact with the agent since s/he does not expect it to notice it anyway. Furthermore, there is empirical evidence that humans tend to avoid computer-controlled agents when navigating through a virtual 3D environment (see a study by Bailenson and colleagues [2] on social group dynamics in virtual realities which seems, however, to be in conflict with the observations by Colburn and colleagues [5] who assume that humans might feel less shy to address an agent).

2 Testbed for Our Research

As a testbed for our research, we used Gamble, a small game of dice (known as Mexicali) where one of the game partners is substituted by a synthetic agent (see Fig. 1). To win the game, it is indispensable to lie to the other players and to catch them lying to you.

The traditional (not computer-based) version of the game is played with two dice that are shaken in a cup. Let's assume player 1 casts the dice. He inspects the dice without permitting the other players to have a look. The cast is interpreted in the following way: the higher digit always represents the first part of the cast. Thus, a 5 and a 2 correspond to a 52. Two equal digits (11, ..., 66) have a higher value than the other casts, the highest cast is a 21. Player 1 has to announce his cast with the constraint that he has to say a higher number than the previous player. For instance, if he casts a 52, but the previous player already announced a 61, player 1 has to say at least 62. Now player 2



Fig. 1. The setting

has to decide whether to believe the other player's claim. In this case, she has to cast next. Otherwise, the dice are shown and if player 1 has lied he has lost this round and has to start a new one. For the experiment, each player was equipped with a PDA which replaced the cup with the dice in the original game.

The implementation of the synthetic game partner is based on the Greta agent system developed by Catherine Pelachaud and colleagues [13,8]. The Greta system is compliant with the MPEG-4 standard which allows to control the facial expressions and body gestures by so-called facial animation parameters (FAPs) and body animation parameters (BAPs).

In the game, Greta tries to mislead the other players by portraying facial expressions that do not correspond to her actual emotional state. For instance, she might express false joy to make her game partners believe that she achieved a high score. Nevertheless, Greta does not lie in a perfect manner, but still reveals her deceptive behaviors by subtle facial cues. Based on Ekman [6], we modeled 32 facial expressions that convey such deceptive cues by combining different degrees of masking with different degrees of asymmetry. For instance, humans involuntarily reveal information regarding a deceptive expression of joy often by an asymmetric smile (see [14] for a more detailed description of our implementation of the deceptive behaviors).

To make the interaction more natural and to enhance the entertaining factor of the Gamble system, we dubbed the animations of the Greta agent with a human voice. Moreover, a full body agent was used and a number of emblematic german gestures were modelled relying on the descriptions in the Berlin dictionary of everyday gestures ("Berliner Lexikon der Alltagsgesten", [4]).

We consider Gamble as a useful testbed for our studies since it allows us to study attentive behaviors to a synthetic agent as well as attentive behaviors to another human player in a multi-party setting. Furthermore, both the synthetic agent and the human players may be in the role of a speaker (when announcing casts or believing/disbelieving

other players) or in the role of an addressee (when listening to announcements of casts or belief statements). Finally, the rules of the game determine who gets the next turn. At least in the case of announcements and belief statements, there is usually no doubt who is addressed by whom.

3 Experimental Method

3.1 Subjects and Design

Subjects were 24 students, all native speakers of German, recruited from the computer science and philosophy faculties at Augsburg University. 12 students from each faculty in their second and third year of study participated, 14 male and 10 female.

As independent variables, we defined the type of interlocutor (ToI) with the levels Human vs. Agent and the user's role in the dialogue (RoU) with the levels Speaker vs. Addressee. Both variables were manipulated between-subjects.

The value of the independent variables depends on the position of the single players. If the subject is standing on the right-hand side in Fig. 1, s/he has to listen to the agent's announcements and to make announcements to the human player on his/her left. If the subject is standing on the left-hand side in Fig. 1, s/he has to listen to the announcements of the human player on his/her right and to make announcements to the agent.

As dependent variables, we defined the length and number of attentive behaviors directed to the conversational partner.

3.2 Procedure

The subjects were randomly divided into 12 teams. At the beginning of the experiment, the subjects were presented with a three minute video of the Gamble system. In addition, they had to participate in a test round to get acquainted with the game, the handling of the PDA and the Greta agent. After the test round, each team played two rounds of 12 minutes. The participants changed positions after the first round so that each participant came to play before and after the agent. We told the subjects that the agent might not be able to conceal her emotions perfectly, but left it open how deceptive behaviours might be detected. Consequently, the subjects had no idea which channel of expression to concentrate on or whether it would be easier to recognize deceptive clues from the behaviors of the agent than from the behaviors of the other human player. To increase interest in the game, the winner was paid five Euros. We videotaped the interactions, and we logged the game progress for the analysis.

3.3 Coding Scheme

The videos are coded for utterances, gaze, role in the game, and laughing. Following [15], gaze is interpreted as indicators for the level of attention. Moreover, we coded who was the current player and because subjects sometimes although rarely addressed the experimenter, a track was introduced for his utterances. Coding is done in Anvil [10] and Fig. 2 gives an impression of the annotation board. It contains the following tracks:

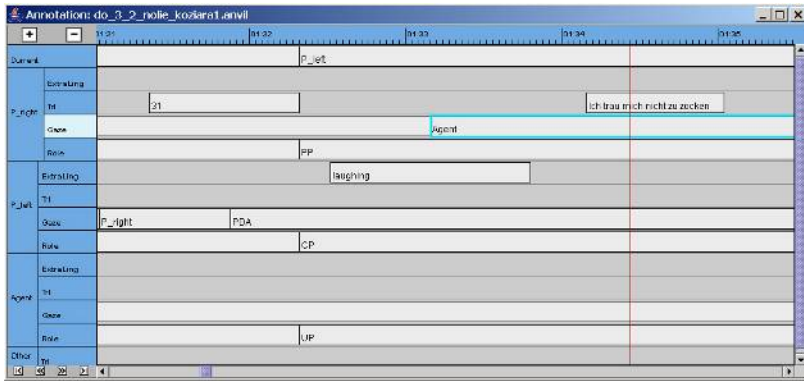


Fig. 2. The annotation board

- *Current*: Indicates the current player, i.e., the player that has to announce her belief/disbelief and that has to cast the dice. Possible elements are Agent, P_right, and P_left.
- *P_right*: Group of tracks for the right player¹ who plays after the agent. The group consists of
 - *ExtraLing*: For annotating if the player is laughing. This information will be used to test automatic recognition of emotions from speech.
 - *Trl*: In this track the utterance of the player is annotated. Utterances are coded per sentence to minimize the coding effort.
 - *Gaze*: The head movements of the player are given in this track. They are interpreted as gaze towards different entities in the environment.² Possible elements are Agent, P_right, P_left, PDA, Camera, and Elsewhere. Coding of gaze behaviors was adopted from Nakano et al. ([12]). A gaze is defined in the following way. The gaze ends and a new one starts in the moment the head starts moving. The direction of the gaze is determined at the end of the head movement.
 - *Role*: This is a secondary track that is bound to the Track *Current*. It specifies the role of the player at the moment in the game. Possible elements are Current, Previous, and Unaffected. Current duplicates the information present in the primary track. Previous indicates that the player is judged by the current player in this turn, and Unaffected indicates that it is the player who is on turn next.
- *P_left*: see P_right
- *Agent*: see P_right
- *Other*
 - *Trl*: The utterances of other people like the experimenter. This track is very rarely used.

¹ The right player seen from the perspective of the coder, not from the agent. This minimizes problems with left/right distinctions, because no cognitive transformations are necessary.

² This method is of course not nearly as accurate as using an eye tracker but has the obvious advantage of unobtrusiveness for the user during the interaction. Due to the spatial setting of the interaction (see Fig. 1), it is moreover necessary to move the head to look at the agent, the other human player, or the PDA thus allowing for a good reliability in coding.

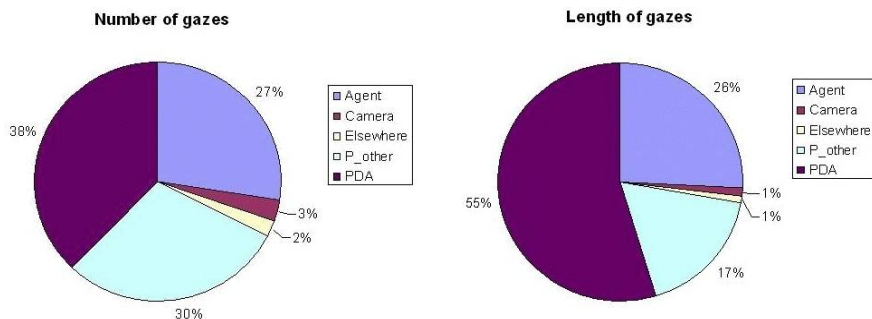


Fig. 3. Number of gazes and length of gazes

Until now, half of the material has been coded by 24 coders. Each video sequence is coded by two people. The interrater reliability has been calculated for gaze and is very good with a kappa value of around 0.9 for each pair. At the moment, the corpus contains 2200 utterances of which 645 are done by the agent, 675 by the right player, and 700 by the left player. Moreover, we have 5398 head movements which are interpreted as gaze behavior of the users.

4 Results of the Analysis

The goal of the annotation work was twofold. On the one hand, we were interested in collecting information to develop an appropriate gaze model for agents in multi-party scenarios. On the other hand, analyzing the users' gaze behavior should reveal to what extent they regard the agent as a real game partner worthy of communication.

Starting with some basic statistics, Figure 3 (left) shows the number of gazes towards each of the given directions. The total number of gazes is 5398. The players looked roughly as often towards the synthetic agent (27%) as towards the other human player (30%). Just considering the number of the gazes, the agent seems to be as attractive as the other player. The fact that people look slightly more often at the PDA (38%) could be attributed to its use as the interface for casting the dice and indicating belief or disbelief.

If we examine instead the length of the gazes towards each of the given directions (Fig. 3 right), this interpretation no longer holds. More than half of the time the players look at the PDA (55%), which seems to bind a lot of their attention. Noteworthy is the fact that players spend considerably more time (1.5) looking at the agent (26%) than looking at the other player (17%). Obviously, the type of interlocutor (human or agent) influences the users' gaze behavior.

The total number of gazes and the length of gazes during the game provide a rough impression of the users' attention towards human and synthetic interlocutors. In addition, we are interested in the question of whether the users' gaze behaviors depends on their role as a speaker or as an addressee. Because Gamble is a strictly round-based game, the utterances can be categorized into three main categories: announcement, belief, and comment. During announcements, the current player announces his cast or

what he pretends to be his cast to the next player who is the addressee of this announcement. The belief category comprises utterances indicating a player's belief or disbelief of an announcement. Hence, the addressee of such an utterance is the previous player who made the announcement that is subject to the speaker's evaluation. All other utterances are categorized as comments which are - strictly speaking - not game-relevant. Among other things, utterances in this category comprise general comments on the game or on the behavior of other players. For the analysis conducted in this paper, comments are disregarded since we are mostly interested in conversational utterances with uniquely determined addressees. In our future work, we will consider comments to study gaze behaviors in situations where the addressee cannot be identified with certainty or where several conversational partners are addressed simultaneously.

Table 1. Gaze behavior of speaker towards addressee and vice versa

User's Role in the Dialogue (RoU)					
Speaker			Addressee		
Interlocutor (ToI)			Interlocutor (ToI)		
Agent	Human	Result	Agent	Human	Result
9.33	8.75	F(1,23)=0.77	31.75	20.17	F(1,23)=23.87

Table 1 compares the gaze behaviors of human interlocutors in the role of a speaker and an addressee for game-relevant utterances. A comparison of the speakers' and addressees' gaze behaviors only makes sense for human interlocutors because the agent is driven by a gaze model (which is not the subject of our investigations). We further distinguish whether their interlocutor is an agent or another human user.

No significant difference was observed in the gaze behavior of the speaker in the two conditions (i) agent (as interlocutor) and (ii) human (as interlocutor). That is people did not apply different gaze behaviors when talking to an agent. Interestingly, similarities in gaze behaviors towards human and artificial addressees could also be observed when humans were lying about their results. In more than 90 % of the cases, people averted the gaze from their game partner when they were lying independently of whether the game partner was human or synthetic. This finding is indeed surprising since it was quite obvious that the agent is not able to read the users' faces.

Turning to the addressee's gaze behavior gives a different picture. Whereas the speaker seems to be uninfluenced by the fact that one of his/her interaction partners is an agent, the addressee's gaze behavior shows a strong significant effect ($F(1,23)=23.97$, $p<0.05$) between the two conditions (i) agent (as interlocutor) and (ii) human (as interlocutor). Being spoken to by the agent grabs the user's attention significantly more than being spoken to by another human interaction partner.

The general patterns of gaze behavior adhere to the literature in that speakers look less at the interlocutor than addressees. Nevertheless, the average gazing time towards the speaker or the addressee is rather low compared to findings by [1]. As noted before, the PDA interface binds a lot of attention. Since the same conditions hold for the human users and the synthetic agents, the experiment should nevertheless enable a fair comparison of human-human and agent-human conversations.

Table 2. Gaze behavior of speaker towards addressee

Gaze Behavior	Agent	Human	Result
Total	9.33	8.75	$F(1,23)=0.77$
begin of utterance	3.25	3.08	$F(1,23)=0.23$
end of utterance	7.33	7.33	$F(1,23)=0.00$

Table 3. Gaze behavior of addressee towards speaker

Gaze Behavior	Agent	Human	Result
Total	31.75	20.17	$F(1,23)=23.87$
begin of utterance	24.58	17.08	$F(1,23)=15.56$
end of utterance	21.5	13.67	$F(1,23)=10.98$

As mentioned in the introduction, some interesting effects can be found at utterance boundaries. Therefore, we analysed the gaze behavior of speaker and addressee at the beginning and end of utterances, again for the two conditions agent and human (see Table 2 and 3). The speaker looks less to the addressee at the beginning and more at the end of an utterances independent on the condition agent or human and thus adheres again to the effects found in the literature [9]. In contrast to the speaker, the addressee tends to look less to the speaker at the end of an utterance. This effect has also been observed for both conditions. Unsurprisingly, there is a significant effect concerning the amount of gaze behaviors in the two conditions for both the beginning of the utterance ($F(1,23)=15.56$) as well as the end of the utterance ($F(1,23)=10.98$). The addressees follow a pattern of gaze behavior that is typical of human-human conversation, but the agent is attracting more attention than the other human user.

5 Discussion

The analysis of the users' gaze behavior in the multiuser setting revealed two effects. In general, users adhere to patterns of gaze behaviors for speaker and addressee that are also reported for dyadic human-human interactions. But they look significantly more often (1.6 times) to the agent when it is talking to them.

One explanation might be that users feel more comfortable to look at a synthetic agent than at a real person - especially when they have not met the other player before. That is to say, it could be less threatening or less embarrassing to concentrate on the agent, a more or less neutral technical artifact than on an unknown human. When looking at our data, it turned out, however, that only one of the six pairs for which the video recordings have already been annotated did not know each other before the game. We therefore consider shyness towards the other human interaction partner as a rather unlikely reason. Furthermore, shyness would not explain why the effect could only be observed in the user-as-addressee condition.

Another explanation for the longer looking time might be the users' hope to find deceptive cues in the agent's face. If we consider the reaction to the other human player

as a baseline, the human addressees were indeed staring at the talking agent. Generally, staring at someone is either a sign of not understanding or interpreted as a threat. Maybe, people need longer to read faces of artificial agents while they restricted themselves to a short glance in the case of human game players.

A more plausible explanation of the user's strong attention towards the agent is the novelty effect of the exceptional conversational partner. None of the participants had encountered an embodied conversational agent in an application yet. All of the participants had already seen agents as manifestations of a new interface metaphor in their courses – and especially the agent employed in Gamble – but they had not interacted themselves with an agent so far. To counter this effect, subjects were shown a three minute video of the interaction during recruiting. To familiarize with the agent and the game, they played for five minutes before the experiment started. Moreover, every couple played two rounds of the game. Thus, when the second round started, every subject had already interacted for 17 Minutes. Even though the participants got some time to familiarize with the agent, the sensation of interacting with a synthetic agent might have persisted for a longer time. Furthermore, the size of the agent that has been projected on the wall might also have increased the user's attention. We suppose that the effect did not occur in the user-as-speaker condition since the agent did not make use of very sophisticated listener behaviors.

An analysis of the address forms employed by the users for the agent leads to interesting observations regarding the relationship between user and agent. Although users were expected to use the PDA interface to interact with the agent when making or responding to announcements, they occasionally addressed the agent directly, for instance, by uttering *Ähh, ich glaub's dir nicht*³. The user directly addresses the agent using the familiar *dir*. Far more frequent are utterances where the users talk about the agent, e.g., *Vielleicht glaubt sie's dir ja*⁴ using the third person singular *sie*. Taking into account utterances containing personal pronouns and disregarding neutral game-relevant utterances, such as *Glaube ich*⁵, 62% of the utterances were classified as talking-about and 38% as talking-to events. Talking about someone who is actually present during the interaction is usually considered as a gross violation of politeness in human face-to-face communication. Such a behavior is, however, typical of conversations involving babies and pets. Bergmann [3] presents categories of utterances found in talks about pets which include explanations of behavior and commenting on behavior. Both categories are also frequently found in talking about the agent which once more stresses the point that users try to make sense of this technical artifact interacting with them.

6 Conclusion

In this paper, we reported on an experiment we conducted to investigate to what extent humans regard synthetic agents as genuine conversational partners. In particular, we analyzed gaze behaviors in human-agent conversations as an important indicator of

³ I don't believe you.

⁴ Perhaps she believes you.

⁵ I believe it.

conversational attention. Unlike earlier studies, we focus on gaze behaviors in multi-party scenarios consisting of human and synthetic interlocutors.

On the one hand, we were able to confirm a number of findings about attentive behaviors in human-human conversation. For instance, our subjects spent more time looking at an individual when listening to it than when talking to it - no matter whether the individual was a human or a synthetic agent. Furthermore, the addressee type (human vs. synthetic) did not have any impact on the duration of the speaker's gaze behaviors towards the addressee. Even though the game was in principle playable without paying any notice to the agent's nonverbal behaviors, the users considered it as worthy of being attended to. Surprisingly was the observation that people avoided gaze contact with the agent when they were lying.

While the users' behaviors in the user-as-speaker condition were consistent with findings for human-human conversation, we noticed differences for the user-as-addressee condition. People spent more time looking at an agent that is addressing them than at a human speaker. Maintaining gaze for an extended period of time is usually considered as rude and impolite. The fact that humans do not conform to social norms of politeness when addressing an agent seems to indicate that they do not regard the agent as an equal conversational partner, but rather as a (somewhat astonishing) artefact that is able to communicate. This attitude towards the agent was also confirmed by the way the users addressed the agent verbally.

In contrast to dyadic agent-user interactions, the multiparty setting allows for meta-communication with the other human interaction partner about what is going on. A rich source for such analysis are the comments which have been disregarded in the current paper but will be examined in the future.

To refine the agent's model of gaze behaviors, our future work will concentrate on an analysis of the human-human interactions in the scenario whereby we will pay special attention to the verbal and situative context. In particular, we are interested in studying the gaze behaviors of people that deliberately convey wrong emotions.

Acknowledgement

The work described in this paper was partially funded by by the EU Network of Excellence Humaine. We are grateful to Catherine Pelachaud, Maurizio Mancini, and Björn Hartmann for supporting our work with the Greta Agent.

References

1. M. Argyle and M. Cook. *Gaze and Mutual Gaze*. Cambridge University Press, Cambridge, 1976.
2. J. N. Bailenson, J. Blasovich, A. C. Beall, and J. M. Loomis. Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin*, 29(7):819–833, 2003.
3. Jörg R. Bergmann. Haustiere als kommunikative Ressourcen. *Soziale Welt: Zeitschrift für sozialwissenschaftliche Forschung und Praxis, Sonderband: Kultur und Alltag*, 8:299–312, 1988.

4. BLAG. Berliner Lexikon der Alltagsgesten. <http://www.ims.uni-stuttgart.de/projekte/nite/BLAG/>, last visited: 09.12.2004.
5. A. Colburn, M. Cohen, and S. Drucker. The role of eye gaze in avatar mediated conversational interfaces, 2000.
6. Paul Ekman. *Telling Lies — Clues to Deceit in the Marketplace, Politics, and Marriage*. Norton and Co. Ltd., New York, 3rd edition, 1992.
7. Maia Garau, Mel Slater, Simon Bee, and Martina Angela Sasse. The impact of eye gaze on communication using humanoid avatars. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 309–316, New York, NY, USA, 2001. ACM Press.
8. B. Hartmann, M. Mancini, and C. Pelachaud. Formational parameters and adaptive prototype instantiation for mpeg-4 compliant gesture synthesis. In *CASA 2002*, pages 111–119, 2002.
9. A. Kendon. Some functions of gaze direction in social interaction. *Acta Psychologica*, 32:1–25, 1967.
10. Michael Kipp. Anvil - a generic annotation tool for multimodal dialogue. In *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech)*, pages 1367–1370, Aalborg, September 2001.
11. Sooha Park Lee, Jeremy B. Badler, and Norman I. Badler. Eyes alive. In *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 637–644, New York, NY, USA, 2002. ACM Press.
12. Y.I. Nakano, G. Reinstein, T. Stocky, and J. Cassell. Towards a model of face-to-face grounding. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2003)*, pages 553–561, 2003.
13. C. Pelachaud and I. Poggi. Subtleties of facial expressions in embodied agents. *Journal of Visualization and Computer Animation*, 31:301–312, 2002.
14. M. Rehm and E. André. Catch me if you can: Exploring lying agents in social settings. In *Proceedings of Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*. ACM Press, 2005.
15. Candace L. Sidner, Cory D. Kidd, Christopher Lee, and Neal Lesh. Where to look: a study of human-robot engagement. In *IUI '04: Proceedings of the 9th international conference on Intelligent user interface*, pages 78–84, New York, NY, USA, 2004. ACM Press.
16. Roel Vertegaal, Robert Slagter, Gerrit van der Veer, and Anton Nijholt. 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*, pages 301–308, New York, NY, USA, 2001. ACM Press.
17. Vinoba Vinayagamoorthy, Maia Garau, Anthony Steed, and Mel Slater. An eye gaze model for dyadic interaction in an immersive virtual environment: Practice and experience. *Comput. Graph. Forum*, 23(1):1–12, 2004.