From chatterbots to natural interaction

Face to face communication with Embodied Conversational Agents.

Rehm, Matthias; André, Elisabeth

Published in:
IEICE Transactions on Information and Systems

Publication date:
2005

Document Version
Accepted author manuscript, peer reviewed version

PAPER

From Chatterbots to Natural Interaction — Face to Face Communication with Embodied Conversational Agents

Matthias REHM[†] *and* **Elisabeth ANDRÉ**[†]**,**

SUMMARY

In this paper, we present a game of dice that combines multi-party communication with a tangible interface. The game has been used as a testbed to study typical conversational behavior patterns in interactions between human users and synthetic agents. In particular, we were interested in the question to what extent the interaction with the agent can be considered as natural. As an evaluation criterion, we propose to investigate whether the communicative behaviors of humans differ when conversing with an agent as opposed to conversing with other humans.

key words: *embodied conversational agents, multi-party communication*

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 (ECA) that employs gestures, mimics and speech to communicate with the human user. While earlier research focused on dyadic communications between a single user and a single agent, more recent research concentrates on multi-party dialogue settings that support reactive as well as proactive user participation. One basic idea is to provide the user with the option of taking an active role in the dialogue if she or he wishes to do so. If not, however, the characters will continue the conversation on their own - maybe encouraging the user to give feedback from time to time.

In this paper, we go one step further and present a game application in which several users interact with a single agent. Such a scenario represents a number of challenges for the artificial agent who has to overhear conversations between humans and potentially engage in different threads of communication. In particular, there is the danger that the humans exclude the agent from the conversation due to its deficient communicative skills. Research on dyadic conversations between humans and agents has shown that an appropriate eye gaze model may positively contribute to the user's engagement in a dialogue. The question arises of whether these findings scale up to multi-party conversations. That is will users still regard an artificial agent as an equal conversational partner worthy of being attended to even if there are human conversational partners around to converse with?

To investigate this question, we developed a multi-party scenario in which several human participants play a small game of dice with an artificial game player. The users are allowed to freely interact with the agent and the other users. Nevertheless, the rules of the game (and the players' motivation to win the game) constrain the interaction in a natural manner. Rather, than asking users for their subjective impression of the agent, we propose to examine whether humans behave similarly when talking to an agent as when talking to a human. In particular, we focus on the users' gaze during the interaction and their verbal utterances.

2. Face to face communication with a single ECA

Early attempts to simulate machine-based dialogues were – due to technical constraints – purely verbal (and text-based) in nature. A system whose concepts of interaction persevere for nearly 40 years now is Weizenbaum's Eliza [1]. The program analyzes the user's input by searching for keywords and employing templates to generate an answer, usually a question or proposal based on the keyword, like *Why are you unhappy* (keyword: unhappy) or *Tell me more about boats* (keyword: boat). The intriguing "trick" of Eliza is to emulate the interaction of certain kinds of psychotherapist (Rogerians) resulting in the astonishing effect that users readily assume to talk to such a person.

Chatterbots are the modern version of the Eliza program, inhabiting a large number of websites and providing customers with a more personalized online-shopping experience 24 hours a day, 7 days a week. In most cases the user can "talk" to the character by typing NL expressions into a text-input widget while the character talks to the user either by voice output or likewise through speech bubbles. Although nowadays web-based chatterbots generally come in the disguise of an embodied agent, most of the time the interaction remains purely text-based enriched perhaps with some gestural output gimmicks. The virtual chat agent Cybelle (www.agentland.com) is an example of this kind.

Embodied conversational agents [2] offer great promise to more natural interaction because of their po-

Manuscript received January 1, 2005.
Manuscript revised January 1, 2005.
Final manuscript received January 1, 2005.

[†]University of Augsburg

Fig. 1 Single User and Multiple Agents (left), Single Agent and Multiple Users (right)

tential to emulate verbal and non-verbal human-human interaction. In general, nonverbal interaction comprises facial expressions, gaze behavior, gestures, and body posture, which all play sometimes distinct, sometimes redundant roles in face to face communication. Most research prototypes of embodied conversational characters aim at the modeling of complex multimodal dialogues, though the focus is usually on the generation of synchronised multimodal expression. Prominent examples include Peedy developed at Microsoft Research [3], the Internet Advisor Cosmo [4], the Steve Agent [5], the real estate agent REA [6], the GRETA Medical Advisor [7], the agent MAX [8] and the animated interface character Smartakus that has been developed in the SmartKom project [9]. Most of these systems rely on sophisticated models for multimodal output generation. For instance, Smartakus incorporates a spoken dialogue subsystem and has a "visual sense" that enables it to recognize and understand pointing gestures of the user.

Summing up, it may be said that the ability of a character to engage with a human in an unconstrained spoken natural language conversation is most desirable, but also very difficult and therefore will remain a great challenge for years even though considerable progress has been made in the last decade in speech recognition, synthesis and spoken dialogue systems. For this reason, we have decided to choose a scenario in which the interactions can be controlled in a natural manner, but without explicitly forbidding the user to engage in free conversations with other users or the agent.

3. Face to face communication with more than one interaction partner

All of the above systems focus on dyadic interactions between one user and one agent. If we turn to communications with more than two interactions partners, we find systems where one user engages in the interaction with several agents. Scenarios with multiple characters bear a number of advantages. First of all, they enrich the repertoire of modalities. For instance, they allow a system to convey certain rhetorical relationships, such as pros and cons, in a more canonical manner (see, for example, [10]). Secondly, the single members of a character team can serve as indices, which help the user to organize the conveyed information. For instance, characters can convey meta-information, such as the origin of information, or they can present information from different points of view, e.g., from the point of view of a businessman or the point of view of a traveller. Furthermore, scenarios with multiple characters allow us to model interpersonal social relationships (see [11], [12]).

A number of approaches to such multiparty conversations have been inspired by research on interactive drama that aims at integrating a user in a scenario - either as an audience member or an active participant. An example includes Avatar Arena [13] where the agents perform a presentation for the user interacting amongst each other. Avatar Arena provides a spatially extended interaction experience by offering several separated agent screens, and by creating the illusion that the agents have cross-screen conversations. An interesting feature of Avatar Arena is the simulation of listener as well as speaker behaviors based on empirical studies of human-human conversations (see left-hand side of Fig. 1). Traum and Rickel [14] have addressed the issue of automatically generated multiparty dialogues in immersive virtual environments. In the context of a military mission rehearsal application, they address dialogue management comprising human-character and character-character dialogues. The characters are based on the Steve architecture which has been enhanced by a multi-modal dialogue model to handle turn taking in such a challenging scenario. The VicTec system (e.g., [15]) realizes a multi-agent learning environment to teach kids strategies against bullying relying on a Forum Theatre metaphor. The user is able to interact with one of the agent and suggest plans of action, that will influence the storyline. In the NICE fairy tale game [16], a user can enter the fairy tale world of H.C. Andersen where she may meet three different types of agents. The helper agent accompanies her through the world, suggesting lines of action. Feature characters have key functions in the plot. To

continue the game, the user has to interact with them. Supporting characters at last are information delivering agents that cannot be engaged in any further interaction.

Hardly any work so far has been conducted on the realization of scenarios with multiple users and synthetic agents. An exception includes the work by Isbister and colleagues [17] who concentrate on social interaction between several humans in a video chat environment which is supported by a so-called Helper Agent. Helper Agent is an animated, dog-faced avatar that tracks audio from two-person conversations and intervenes if it detects longer silences.

In contrast to [17], we focus on a game scenario in which the agent does not appear in the role of a moderator, but takes on a similar role as the human users (see right-hand side of Fig. 1). In our case, the dialogue flow is controlled by the rules of the game. As a consequence, we do not rely on dramaturgical principles to structure the conversation.

4. The role of gaze behaviors in dyadic and multi-party conversations

In this section, we focus on eye gaze behaviors as one of the most important signals to show engagement in a conversation. 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. 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 [19] show that people look nearly twice as much while listening (75%) than while speaking (41%).

For dyadic interactions between an ECA and a human, a positive effect of natural gaze behavior was found. Nakano and colleagues [20] developed a model of grounding for the kiosk agent Mack agent 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 [21] 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.

Little is known of the effects when we move form a dyadic agent-user interaction towards a situation where the agent interacts with more than one user. In a dyadic interaction, the user can concentrate completely on the task and the agent who is the sole interlocutor. Increasing the number of human communication participants, there is suddenly a choice between a natural communication partner with sophisticated communication skills and an artificial communication partner with deficient communication skills. Thus, we cannot trust that the results of the dyadic studies scale up 1:1 to such an extended setting.

While the work above concentrated on the effect of agent's gaze behaviors on the user's interaction behaviors, we focus on the question of whether and how the users' gaze behaviors change when they are addressing or attended by another human interlocutor as opposed to artificial interlocutors.

5. Multiparty face to face communication in Gamble

In the remaining part of this article we focus on Gamble, a system that allows for the investigation of multiparty interactions (see Fig. 2). In Gamble, two users play a simple game of dice (also known as Mexicali) with an embodied conversational agent. To win the game it is indispensable to lie to the other players and to catch them lying to you. The game is played with two dice and a cup.

Let's assume player 1 is on turn. He casts the dice and then 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 the dice show a 52, but the previous player already announced a 61, player 1 has to say at least 62. Now player 2 has to decide whether to believe the other player's claim. In this case, he 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.

Although the rules of the game are very simple, complex communicative behaviors emerge from these simple rules. Blaming another player for an attempted deceit or getting away with such an attempt e.g., creates highly emotional situations that trigger rich verbal and nonverbal interactions allowing us to use Gamble as a test bed for investigating multiparty communications.

Gamble is based on the Greta agent system developed by Catherine Pelachaud and colleagues [7]. It is compliant with the MPEG-4 standard which allows

to control facial expressions and body gestures by so-called facial animation parameters (FAPs) and body animation parameters (BAPs).

5.1 Handling the complexity

The choice of this game domain helps us to manage some of the inherent complexities of multi-party interactions. The game is turn-based which means that at each moment in time exactly one player legally holds the floor and it is unmistakably clear who this player is and whom he is speaking to. In this way, the process of turn-taking is controlled in a natural manner by the game progress. Verbal and non-verbal behavior can be classified into three different categories: announcements, beliefs, and comments. During announcements, the current player announces her cast or what she pretends to be her cast to the next player who is the addressee of this announcement. The belief category comprises communicative acts 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 communication attempts are categorized as comments which are – strictly speaking – not game relevant. Gamble is a small game of dice played at a table. Similar to Avatar Arena, this setting provides the user with a spatially extended interaction experience. By projecting the agent on a screen at the end of the table, we convey the impression that it is sitting together with the other players at the table. To allow for a more natural simulation of the traditional game of dice, the user does not interact via a keyboard, but uses a tangible interface to toss the dice and communicates via voice with the other players and the agent.

5.2 Multimodal input behavior

PDA interaction In the preliminary version of Gamble, users were restricted to a handheld device (PDA) as an interface to the system (see right-hand side of Fig. 1). The PDA is used to throw the dice by clicking on a displayed cup of dice, to enter an announcement that will be sent to the game server, and to indicate belief or disbelief by selecting yes/no buttons. The players are forced to communicate their input also verbally because the information is not displayed on the PDA of the other player. Of course, this is not a natural way to play a game of dice. Indeed, we noticed that the PDA took away a lot of attention from the players which is why we developed a specialized input device for the application.

CamCup For this specific game of dice, a cup is used to cast the dice in order to allow the players to hide their casts. We developed the CamCup as a tangible interface for this application. The CamCup (see Fig. 2) contains an USB-camera in order to recognize the dice cast. The game server needs this information about the actual result of the cast to decide on the game progress in case the next player does not believe the current player's announcement. If the current player has indeed tried to deceive the next player, i.e., he has announced a result that was higher than his actual cast, he has lost the round. Otherwise the next player has lost the round. Either way, the loser has to start a new round. The CamCup is a natural device to play a game of dice because it is just an elongated cup of dice which is used as every other cup.

Speech recognition Apart from casting the dice, the players have to announce their results and evaluate the other players announcements. To capture the players' verbal input, a microphone is set between the players. The utterances are analyzed by the speaker-independent speech recognition system ESMERALDA [22]. It was trained to recognize all possible casts and a few variations of "yes" and "no" like "I believe you" or "Never". Synchronizing the different input modalities is simple because the players have to cast and inspect the dice first before an announcement can be registered. Thus, any utterance before and during casting the dice can be ignored if an announcement has to be detected. A belief statement on the other hand, can only be made after an announcement and generally precedes casting the dice.

5.3 Multimodal output behavior

Verbal behavior To render the agent's voice as natural as possible, Gamble uses animation sequences that were dubbed by a female German native speaker and that are chosen from a large database at runtime. This has the disadvantage of some repetition during long game sessions. At the moment, a German speech synthesizer is integrated into the system to allow for a dynamic and situative generation of utterances. The agent's verbal behaviors comprise all three categories of utterances, i.e., also comments. Comments are reactions to false accusations (e.g. *Immer dieses Misstrauen*[†]), winning (e.g. *Jetzt musst du was trinken*[††]) or loosing (e.g. *Da kann man wohl nichts machen*[†††]) a round as well as comments on its own casts (e.g. *Was mach ich jetzt nur*[††††]).

Gaze Compared to dyadic conversations, we know little about gaze behavior in multiparty interactions. One of the few studies has been conducted by Vertegaal and colleagues [23] who investigated the gaze behavior in a four-party interaction. Subjects looked about

[†]Always mistrusting me
[††]Now you have to drink something
[†††]Nothing to be done
[††††]What shall I do now

Fig. 2 The Gamble system. The preliminary version allowed for interactions by the use of handheld devices. The current version employs a camera mounted cup of dice and a speech recognition system to capture the users input.

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. [24] or Nakano et al. [20], they conclude that gaze is an excellent predictor of conversational attention in multiparty conversations. Vertegaal et al. also showed that (i) People look more at the person they speak or listen to than at others, (ii) 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 she not been addressed, (iii) Speakers compensate for divided visual attention by increasing the total amount of their gazes, and (iv) Listeners gaze more than speakers (1.6 times). To create a gaze model for the agent in the Gamble scenario, we rely on the studies presented above, but will also exploit the results of our own analysis of user gaze behaviors while interacting with another human in Gamble (see Section 6). Since Gamble is a turn-based game, it is usually obvious who is talking to whom. If the speech recognizer identifies off-talk, it is rather likely that a conversation between the hman participants has started. In this way the agent is at least to some extent able to exhibit natural eye gaze behaviors without tracking where the participants are looking at.

Gestures Following the *Berlin Lexicon of German Everyday Gestures* [25], 30 different gestures were specified for the use in Gamble. The rationale for choosing those gestures was threefold. First, they are well documented including the use, as well as the form and the meaning of the gestures. Second, they are clearly identifiable by German native speakers. 90.5% of the generated gestures are correctly classified by subjects. Third, at least half of the documented gestures have a clearly emotion related meaning, e.g., *Waving a hand in front of one's eyes* thereby indicating that something is stupid, *Indicating to one's wrist* which can be interpreted as hurry up, or *Holding the hand as an extension of the nose*, a clear sign of gloating. This makes them suitable for the use in the game scenario where highly emotional situations arise that ask for the display of appropriate reactions.

In the communicate process, the agent's gestures thus reveal its emotional appraisal of the situation allowing the other participants to form an impression of the agent's emotional state.

Facial Expressions Facial expressions are defined by moving the FAPs of the face model. The original face library of the Greta system defines FAP movements for the display of 48 different emotions. In Gamble it is indispensable to lie to the other participants if you want to win the game. 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. In fact, her emotional state is more that of worry whether she succeds with the bluff or not. Thus, the worried facial expression may leak through the nonchalant smile rendering it not as natural as would be expected of a genuine smile. According to Ekman [26], this is called a masking smile. Another sign of a conscious facial display is asymmetry, e.g., a forced smile may result in more smiling action on one side of the face. Based on Ekman [26], we modeled 32 facial expressions that convey such deceptive cues by combining different degrees of masking with different degrees of asymmetry (see [27] for a more detailed description of our implementation of the deceptive behaviors).

Like gestures, facial expressions allow the agent to reveal its emotional appraisal of a communicative situation. More crucially, they may indicate a deceptive move to the user, an information that will never be revealed by the agent's verbal channel.

Emotions The emotional model influences the agent's decision and behavior selection process. In the communicative process, the agent's emotions are conveyed to the user by facial expressions as well as by appropriate gestures, e.g., tapping her finger against her forehead which indicates in German that someone is nuts.

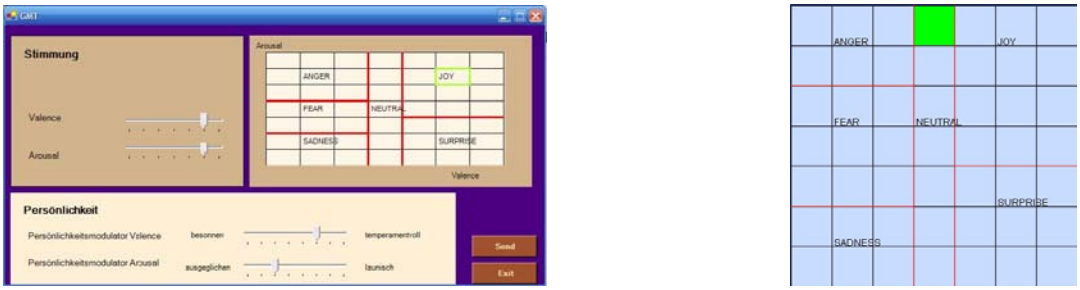

Fig. 3 Setting the agent's initial emotion (left, above) and its personality (left, below) and visualising the agent's emotional state (right).

The agent's emotional state is influenced by its game success and by its personality traits. Catching another player lying, getting away with a lie, or being falsely accused of a lie and thus winning the round constitute a positive emotional influence. Falsely accusing another player or being caught lying on the other hand constitute a negative emotional influence. The emotional model is dimensional in nature (e.g., [28]) with one dimension denoting the arousal of the accompanying emotion and the other dimension denoting its valence on a positive/negative axis. The agent's initial emotional state as well as its personality traits can be set before starting the game (see left-hand side of Fig. 3). Instead of using a sophisticated personality model like the big five (e.g., [29]), we take the dimensional model into account directly. The user can determine modulator values for valence and arousal. These modulators allow the agent different appraisals of a given situation. A high modulator value for valence is interpreted as a highly emotion driven decision process changing fast between positive and negative evaluation of a situation whereas a low modulator value results in a more rational decision. The arousal modulator on the other hand determines how capricious the agent reacts. A high value of the arousal modulator results in a fast increase of the arousal level in a given situation whereas a low value slows the increase down making the agent more phlegmatic. The agent's emotional state can be monitored (see right-hand side of Fig. 3) but is visible to the user only by the agent's facial expression, its gestures and its behavior. A very happy agent e.g., will comment if falsely accused by a gloating gesture towards the loser (*Holding the hand as an extension of the nose*) and an accompanying utterance like *Du hast verloren*[†].

5.4 Behavior control

Before generating the animations, the appropriate, i.e., context- and situation-specific, verbal and non-verbal behaviors have to be decided on. A Bayesian network is deployed for this reasoning process. Depending on the evidence available, the network calculates probabilities for possible actions. A turn in the game can roughly be divided into two phases: rating and announcing. First, the announcement of the previous player has to be rated. This decision is based upon (i) the agent's current emotional state, (ii) the probability of the announcement, (iii) the number of times that the previous player was caught lying. If the agent believes the previous player or has falsely accused him of lying, it has to cast the dice next and announce a result. The announcement is based upon (i) the agent's cast, (ii) the probability of the necessary announcement, (iii) the number of times the agent was already caught lying, (iv) the agent's emotional state.

The information about the actual cast and the announcement are sent to the game server. Output of the behavior module for the animation generation is the result that will be announced by the agent, the emotional state of the agent in terms of valence and arousal, and the current ability to mask a necessary lie. This ability depends in our model on the arousal and valence value of the emotional state, on the probability for the announcement of the previous player, and on the probability of the agent's own announcement.

6. Empirical Evaluation

Instead of presenting the user with a questionnaire to acquire information regarding their subjective impression of the game agent, we decided to perform an objective evaluation of the user's level of engagement in the game by analysing his or her behaviors during the game. In particular, we investigated whether the user's verbal and non-verbal behaviors changed when addressing or attending an agent as opposed to addressing or attending another human interlocutor.

In order to determine the users' level of engagement, we analyzed the recordings of game interactions of two human players with the Gamble system according to the gaze behavior they exhibit. Subjects were 24 students, all native speakers of German, recruited from the computer science and philosophy faculties at Augsburg University. The subjects were randomly divided into 12 teams. At the beginning of the experiment, the

[†]You have lost

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 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. 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.

6.1 Analyzing the user's gaze behaviors

Because the gaze behaviors of speakers and listeners differ, our investigation focused on two questions: (i) Does the type of addressee (agent or human) influence the users' gaze behaviors? and (ii) Does the type of speaker (agent or human) influence their gaze behavior?

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 type of the addressee (agent or human) did not have any significant impact on the duration of the speaker's gaze behaviors towards the addressee. Even though the game can be played without paying any attention to the agent's nonverbal communicative behaviors, the users attended to it. Surprisingly, on the other hand, users spent more time looking at an agent that was addressing them than at a human speaker ($F(1,23)=23.97$, $p<0.05$). 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 attending to 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.

In addition, logging the game progress allowed us to gain insight in the gaze behavior of the users while they were lying about their result. 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.

6.2 Analyzing the user's verbal behaviors

An analysis of the address forms employed by the users for the agent led to interesting observations regarding the relationship between user and agent. Although users mostly used neutral game-relevant utterances when interacting with the agent like *Glaube ich*[†], they occasionally addressed the agent directly using the familiar *dir*, for instance, by uttering *Ähh, ich glaub's dir nicht*[††]. Far more frequent are utterances where the users talk about the agent, e.g., *Vielleicht glaubt sie's dir ja*[†††] using the third person singular *sie*. Only utterances containing personal pronouns were taken into account while neutral game-relevant utterances, such as *Glaube ich* were disregarded. From the remaining utterances, 62% were classified as talking-about and 38% 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 [30]. Our experiment seems to indicate that people feel attracted by the agent's expressiveness, but nevertheless do not regard it as an equal conversational partner.

7. Conclusions

In this paper, we presented a multi-party scenario in which several users converse with an embodied conversational agent who takes on the role of a game partner in a small game of dice. The users interact with the game application via a tangible interface and spoken utterances. To abstract from the inherent problems of unconstrained natural language understanding, we have chosen a scenario with controlled interactions. Furthermore, we decided to implement a turn-based game in order to avoid misconceptions regarding the assignment of turns. The specific characteristics of the scenario helped us to mitigate some of the deficiencies of the agent, but still allowed us to explore typical characteristics of a multi-party scenario without any manual intervention. An empirical evaluation of the system revealed that users exhibit communicative behaviors towards the agent that ressemble communicative behaviors towards human conversational partners. For instance, even though it was quite obvious to the users that the agent is not able to read their faces, they avoided eye contact with the agent when they were lying. On the other hand, we also noticed some pecularities regarding their behaviors towards the agent that seem to indicate that they do not accept the agent as an equal conversational partner.

[†]I believe it

[††]I don't believe you

[†††]Perhaps she believes you

References

[1] J. Weizenbaum, "ELIZA - A Computer Program for the Study of Natural Language Communication between Man and Machine," Communications of the Association for Computing Machinery, no.9, pp.36–45, 1966.

[2] J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, eds., Embodied conversational agents, MIT Press, Cambridge, MA, 2000.

[3] G. Ball, D. Ling, D. Kurlander, J. Miller, D. Pugh, T. Skelly, A. Stankosky, D. Thiel, M.V. Dantzich, and T. Wax, "Lifelike computer characters: the persona project

at microsoft," pp.191–222, 1997.

[4] J.C. Lester, J.L. Voerman, S.G. Towns, and C.B. Callaway, "Deictic believability: Coordinated gesture, locomotion, and speech in lifelike pedagogical agents.," Applied Artificial Intelligence, vol.13, no.4-5, pp.383–414, 1999.

[5] J. Rickel and W. Johnson., "Animated agents for procedural training in virtual reality: Perception, cognition and motor control," Applied Artificial Intelligence, no.13, pp.415–448, 1999.

[6] J. Cassell, T. Bickmore, L. Campbell, H. Vilhjálmsson, and H. Yan, "Human conversation as a system framework: designing embodied conversational agents," pp.29–63, 2000.

[7] C. Pelachaud, B. De Carolis, F. de Rosis, and I. Poggi, "Embodied contextual agent in information delivering application," Proceedings of the first international joint conference on Autonomous agents and multiagent systems, pp.758–765, ACM Press, 2002.

[8] S. Kopp, B. Jung, N. Lessmann, and I. Wachsmuth, "Max — A Multimodal Assistant in Virtual Reality Construction," KI – Künstliche Intelligenz, no.4, pp.11–17, 2003.

[9] N. Reithinger, J. Alexandersson, T. Becker, A. Blocher, R. Engel, M. Löckelt, J. Müller, N. Pfleger, P. Poller, M. Streit, and V. Tschernomas, "Smartkom: adaptive and flexible multimodal access to multiple applications," ICMI '03: Proceedings of the 5th international conference on Multimodal interfaces, New York, NY, USA, pp.101–108, ACM Press, 2003.

[10] T. Rist, S. Baldes, P. Gebhard, M. Kipp, M. Klesen, P. Rist, and M. Schmitt, "CrossTalk: An Interactive Installation with Animated Presentation Agents," Proceedings of the 2nd Int. Conf. on Computational Semiotics for Games and New Media, Cosign, ed. E. Andr¨e, A. Clarke, C. Fencott, C. Lindley, G. Mitchell, and F. Nack, pp.61–67, 2002.

[11] H. Prendinger and M. Ishizuka, "Social role awareness in animated agents," AGENTS '01: Proceedings of the fifth international conference on Autonomous agents, New York, NY, USA, pp.270–277, ACM Press, 2001.

[12] D.V. Pynadath and S. Marsella, "PsychSim: Modeling Theory of Mind with Decision-Theoretic Agents," Proceedings of the Nineteenth IJCAI, Morgan Kaufman Publishers Inc., 2005.

[13] T. Rist, M. Schmitt, C. Pelachaud, and M. Bilvi, "Towards a Simulation of Conversations with Expressive Embodied Speakers and Listeners," CASA, pp.5–10, 2003.

[14] D. Traum and J. Rickel, "Embodied Agents for Multiparty Dialogue in Immersive Virtual Worlds," Proceedings of the first international joint conference on Autonomous agents and multiagent systems, New York, pp.766–773, ACM Press, 2002.

[15] A. Paiva, J. Dias, D. Sobral, R. Aylett, S. Woods, and C. Zoll, "Caring for agents that care: Building empathic relations with synthetic agents," Proceedings of AAMAS 03, 2003.

[16] J. Gustafson, L. Bell, J. Boye, A. Lindstrm, and M. Wiren, "The NICE Fairy-tale Game System," Proceedings of SIGdial 04, 2004.

[17] K. Isbister, H. Nakanishi, T. Ishida, and C. Nass, "Helper agent: designing an assistant for human-human interaction in a virtual meeting space," CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY, USA, pp.57–64, ACM Press, 2000.

[18] A. Kendon, "Some functions of gaze direction in social interaction," Acta Psychologica, no.32, pp.1–25, 1967.

[19] M. Argyle and M. Cook, Gaze and mutual gaze, Cambridge University Press, Cambridge, 1976.

[20] Y.I. Nakano, G. Reinstein, T. Stocky, and J. Cassell, "Towards a Model of Face-to-face Grounding," Proceedings of the Association for Computational Linguistics, Sapporo, Japan, July 1–12 2003.

[21] C.L. Sidner, C.D. Kidd, C. Lee, and N. Lesh, "Where to look: a study of human-robot engagement," IUI '04: Proceedings of the 9th international conference on Intelligent user interface, New York, NY, USA, pp.78–84, ACM Press, 2004.

[22] G.A. Fink, "Developing HMM-based recognizers with Esmeralda," in Lecture notes in artificial intelligence, ed. V. Matoušek, P. Mautner, J. Ocelíková, and P. Sojka, pp.229–234, Springer, Berlin, Heidelberg, 1999.

[23] R. Vertegaal, R. Slagter, G. van der Veer, and A. Nijholt, "Eye Gaze Patterns in Conversations: There is More to Conversational Agents Than Meets the Eyes," Proceedings of SIGCHI 2001, Seattle, WA, 2001.

[24] C.L. Sidner, C.D. Kidd, C. Lee, and N. Lesh, "Where to look: a study of human-robot engagement," Proc. of the 9th Int. Conf. on Intelligent User Interface, pp.78–84, 2004.

[25] BLAG, "Berliner lexikon der alltagsgesten," http://www.ims.uni-stuttgart.de/projekte/nite/BLAG/, last visited: 22.03.2005.

[26] P. Ekman, Telling Lies — Clues to Deceit in the Marketplace, Politics, and Marriage, 3rd ed., Norton and Co. Ltd., New York, 1992.

[27] M. Rehm and E. André, "Catch me if you can — Exploring lying agents in social settings," Proceedings of AAMAS 2005, 2005.

[28] P.J. Lang, "The Emotion Probe: Studies of Motivation and Attention," American Psychologist, vol.50, no.5, pp.372–385, 2002.

[29] O.P. John, "The "Big Five" factor taxonomy: Dimensions of personality in the natural language and in questionnaires," in Handbook of personality: Theory and research, ed. L.A. Pervin, pp.66–100, Guilford, New York, 1990.

[30] J.R. Bergmann, "Haustiere als kommunikative Ressourcen," Soziale Welt: Zeitschrift für sozialwissenschaftliche Forschung und Praxis, Sonderband: Kultur und Alltag, vol.8, pp.299–312, 1988.

Dr. Matthias Rehm He is senior researcher at the Lab for Multimedia Concepts and Applications, University of Augsburg, Germany. His research interests comprise embodied conversational agents, modelling social behavior, multimodal as well as emotional interactions.

Prof. Dr. Elisabeth André She is a full professor for Computer Science at the University of Augsburg and head of the Lab for Multimedia Concepts and Applications. Her research interests include multimodal user interfaces, affective computing and embodied conversational agents.