Chapter 1

CONTROLLING THE GAZE OF CONVERSATIONAL AGENTS

Dirk Heylen, Ivo van Es, Anton Nijholt and Betsy van Dijk
University of Twente, the Netherlands
{heylen, es, anijholt, bvdijk}@cs.utwente.nl

Abstract We report on a pilot experiment that investigated the effects of different eye gaze behaviours of a cartoon-like talking face on the quality of human-agent dialogues. We compared a version of the talking face that roughly implements some patterns of human-like behaviour with two other versions. In one of the other versions the shifts in gaze were kept minimal and in the other version the shifts would occur randomly. The talking face has a number of restrictions. There is no speech recognition, so questions and replies have to be typed in by the users of the systems. Despite this restriction we found that participants that conversed with the agent that behaved according to the human-like patterns appreciated the agent better than participants that conversed with the other agents. Conversations with the optimal version also proceeded more efficiently. Participants needed less time to complete their task.

Keywords: Gaze, embodied conversational agents, human computer interaction.

1. Introduction

Research on embodied conversational agents is carried out in order to improve models and implementations simulating aspects of human-like conversational behaviour as best as possible. Depending on the application or precise research aims, one might strive for the synthetic characters that one is building to be believable, trustworthy, likeable, human- and life-like. This involves, amongst other things, having the character display the appropriate signs of a changing mood, a recognisable personality and a rich emotional life. The actions that have to be carried out by agents in dialogue situations include the obvious language understanding and generation tasks: knowing how to carry out a conversation and all the types of conversational acts this involves (openings, greetings, closings, repairs, asking a question, acknowledging, back-channelling, etc.) and also using all the different modalities, including body-language (posture, gesture, and facial expressions).

Although embodied conversational agents are still far from perfect, some agents have already been developed that can perform quite a few of the functions that were listed above to a reasonable extent and that can be useful in practical applications (tutoring, for instance). The Cassell et al. [2000] collection provides a good overview of such systems.

In our research laboratory we started to develop spoken dialogue systems some years ago. We built an interface to a database containing information on performances in the local theatres. Through natural language dialogue, people could obtain information about performances and order tickets. A second step involved reconstructing one of the theatres in 3D using VRML and design a virtual human, Karin, that embodies this dialogue system (Figure 1.1).

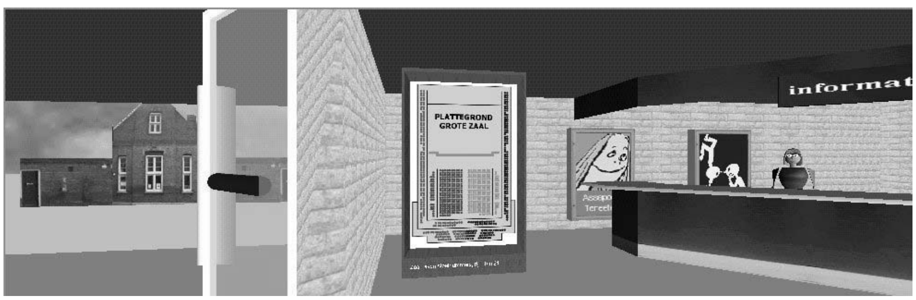

Figure 1.1. The Virtual Music Center.

We first focused the attention on several aspects of the multi-modal presentation of information [Nijholt and Hulstijn, 2000]. We combined presentation of the information through the dialogue system with traditional desktop ways of presentation through tables, pop-up menus and we combined natural language interaction with keyboard and mouse input. We wanted our basic version to be web-accessible which, for reasons of efficiency, forced us at that time to leave out the speech recognition interface from this version.

The dialogue agent, Karin, is placed behind an information desk. In the situation when visitors enter the virtual environment of the Virtual Music Center and approach Karin she will start to speak and ask whether there is anything she can do for the visitor. The browser screen looks as in Figure 1.2 with one part displaying the 3D environment, another part showing the dialogue window to the right and, below these two, a table presenting information about the performances as a result of the user queries. Visitors have to type in their dialogue part. Karin's answers also appear in the dialogue window but will also be pronounced by a text-to-speech system.

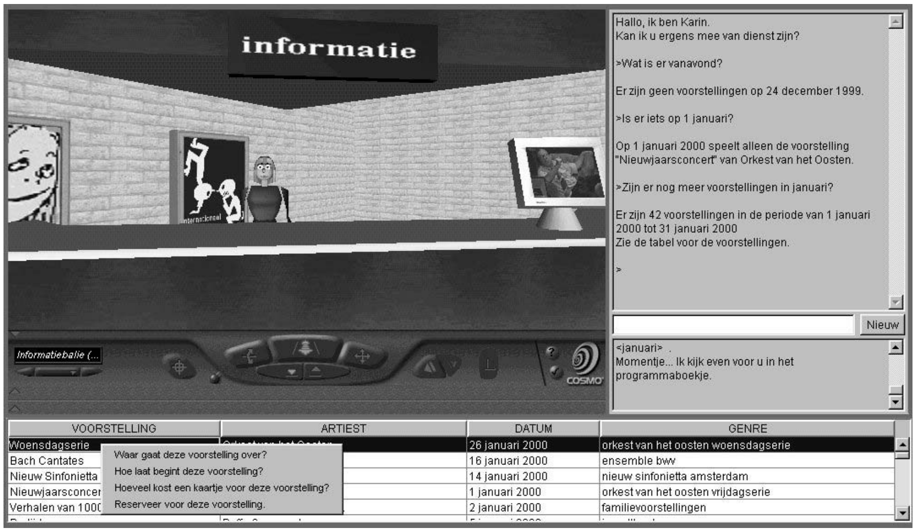

Figure 1.2. Karin in the virtual environment.

We have moved on to implement other types of embodied conversational agents that are designed to carry out other tasks like navigating the user through the virtual environment or agents that act as tutors. Besides the work we did on building other types of agents we have also tried to explore in more depth different cognitive and affective models of agents, including symbolic BDI models as well as neural network models. We have also worked on extending their communicative skills. Current work, as summarised in [Heylen et al., 2001], is concerned with several aspects of non-verbal behaviour including facial expressions, posture and gesture. This chapter deals with one kind of non-verbal behaviour: gaze.

Gaze has been shown to serve a number of functions in human-human interaction [Kendon, 1990]. On a meta-conversational level, gaze helps to regulate the flow of conversation and plays an important role in ensuring smooth turn-taking behaviour. Speakers, for instance, have the tendency to gaze away from listeners at potential turn-taking positions when they want to keep on talking. Listeners show continued attention when gazing at the speaker. On an interpersonal level, duration and types of gaze communicate the nature of the relationship between the interlocutors.

This made us curious about our own situation with the agent Karin. We wondered whether implementing some kind of human-like rules for gaze behaviour would have any effects. We therefore set up an experiment. Although people are talking to an agent in a somewhat unnatural way – they are typing in their input, for instance – previous research had shown that people are sensitive to the gaze of such agents.

In the next section of this chapter we will discuss some aspects of the function of gaze in face-to-face conversations between humans and in mediated forms of communication. Next we describe the experiment we did with our embodied agent Karin and discuss the results.

2. Functions of Gaze

2.1 General

The function of gaze in human-human, face-to-face dialogues has been studied quite extensively, see [Argyle and Cook, 1976] for an overview). The way speakers and hearers seek or avoid mutual eye contact, the function of looking to or away from the interlocutor, the timing of this behaviour in relation to aspects of discourse and information structure have all been investigated in great detail and certain typical patterns have been found to occur. In these investigations a lot of parameters like age, gender, personality traits, and aspects of interpersonal relationships like friendship or dominance, and the nature of the setting in which the conversation takes place have been considered.

In trying to build life-like and human-like software agents that act as talking heads which humans can interact with as if they were talking face-to-face with another human, one is also led to consider the way the agents look away and towards the human interlocutor. This has been the concern of several researchers on embodied conversational agents. It is also related to work on forms of mediated human-human communication as in teleconferencing systems that make use of avatars, for instance. Previous research was mostly concerned with trying to describe an accurate computational model of gaze behaviour. A few evaluations of the effects of gaze on the quality of interactions in mediated conversation (mostly using avatars instead of autonomous agents) have been carried out by Vertegaal [1999], Garau et al. [2001], Colburn et al. [2000] and Thórisson and Cassell [1996], amongst others. These papers have shown that improving the gaze behaviour of agents or avatars in human-agent or human-avatar communication has noticeable effects on the way communication proceeds.

2.2 Human to Human

The amount of eye contact in a human-human encounter varies widely. Some of the sources of this variation as well as some typical patterns that occur have been identified. Women, for instance, are found to engage in eye contact more than men. Cultural differences account for part of the variation as well.

Argyle and Cook [1976] provide an extensive overview of these investigations. We will summarise some of the major findings here.

When people in a conversation like each other or are cooperating there is more eye contact. When personal or cognitively demanding topics are discussed eye contact is avoided. Stressing the fact that the following figures are only averages and that wide variation is found, Argyle [1993] provides the following statistics on the percentage of time people look at one another in dyadic (two-person) conversations. About 75% of the time that they are listening coincides with gazing at the speaker. People that are talking look less of the time at the listener (40%). During a complete interaction there is eye-contact (mutual gaze) only 30% of the time.

Among the common subjective interpretations of eye contact have been found friendship, sexual attraction, hate and a struggle for dominance. This list shows that the same behaviour can result in opposite valuations. The precise quality depends on the circumstances. The simplest inference a person can draw from the situation where someone is looking at him is that the other is paying attention. In a face-to-face conversation this is appreciated positively. However, in public places extensive gazing at strangers may be felt to be impolite or even threatening.

There are individual differences in the amount and type of gaze depending, also, on personality traits. "Gaze levels are also higher in those who are extroverted, dominant or assertive, and socially skilled. Perception of eyes leads to arousal, and to avoidance after a certain period of exposure. The finding that extraverts (who have a low level of arousal) look in general more than introverts is consistent with this hypothesis" [Argyle and Cook, 1976, page 21]. People who look more tend to be perceived more favourably, other things being equal, and in particular as competent, friendly, credible, assertive and socially skilled [Kleinke, 1987].

Besides these more psychological or emotional signal functions of gaze, looking to the conversational partner also plays an important part in regulating the interaction. The patterns in turn taking behaviour and the relation to (mutual) gaze have been the subject of several investigations. Studying the patterns in gaze and turn-taking behaviour, Kendon [1990] was one of the first to look with some detail at how gaze behaviour operates in dyadic conversations. He distinguishes between two important functions of an individual's perceptual activity in social interaction. By looking or not looking, a person can control the degree of monitoring his interlocutor and this choice can also have regulatory and expressive functions.

Argyle and Dean [1972] report that in all investigations where this has been studied it has been found that there is more eye contact when the subject is listening than when he is speaking (cf. Table 1.1). Furthermore people look up at the end of their turn and/or at the end of phrases and look away at the start of (long) utterances, not necessarily resulting in mutual gaze or eye contact. The patterns in gaze behaviour are explained by a combination of principles. Speakers that start longer utterances tend to look away to concentrate on what they are saying, avoiding distraction, and to signal that they are taking the floor and do not want to be interrupted. At the end of a turn, speakers tend to look up to monitor the hearer's reaction and to offer the floor.

Table 1.1. Percentages of Gaze in Dyadic Conversations.

Individual gaze	60 %
While listening	75 %
While talking	40 %
Eye-contact	30 %

In [Cassell et al., 1999], the relation between gaze, turn-taking, and information structure is investigated in more detail. The empirical analysis shows the general pattern of looking away and looking towards the hearer at turn-switching positions. The main finding reported in this chapter, is that if the beginning of a turn starts with the thematic part (the part that links the utterance with previously uttered or contextualised information), then the speaker will always look away and when the end of the turn coincides with a rhematic part (that provides new information), then the speaker will always look towards the listener at the beginning of the rhematic part. In general, beginnings of themes and beginnings of rhemes are important places where looking away and looking towards movements occur. From these observations on the gaze behaviour one could derive some prognoses with respect to the effects of the design of gaze behaviour for an embodied conversational agent. The amount and type of gaze will influence how the character of the agent will be perceived. The patterns of gaze in relation to the discourse and information structure may lead to more or less efficient conversations. However, these are prognoses based on findings about human-human, face-to-face interaction. In the next subsection we look at some studies of gaze behaviour in mediated conversations between human interlocutors and in conversations between humans and agents.

2.3 Mediated Conversations

Several researchers have investigated the effects of implementing gaze behaviour in conversational agents or in other forms of mediated conversation between humans. In videoconferencing for instance, avatars may be used to represent the users.

Vertegaal [1999] describes the GAZE groupware system in which participants are represented by simple avatars. Eye-tracking of the participants informs the direction in which the avatars look at each other on the screen so that the avatars mimick the gazing behaviour of the participants, see also [Vertegaal et al., 2001]. They have shown, in experiments, that this improves such videoconferencing discussions in several ways.

Garau et al. [2001] describe an experiment with dyadic conversation between humans in four mediated conditions: video, audio-only, random-gaze avatars and informed gaze avatars. In the latter case, gaze was related to conversational flow. The experiment showed that the random-gaze avatar did not improve on audio-only communication, whereas the informed gaze-avatar significantly outperformed audio-only on a number of response measures.

Colburn et al. [2000] also describe some experiments in conversations between humans and avatars in a video-conferencing context. One of the questions they asked was whether users that interact with an avatar will act in ways that resemble human-human interaction or whether the knowledge that they are talking to an artificial agent counteracts natural reactions. In one experiment they changed the gaze behaviour of avatars during a conversation. It appeared from this and similar experiments that participants, though not consciously aware of the differences in the avatar's gaze behaviour, still react differently (subliminally).

In the context of embodied conversational agents, rules for gaze behaviour of agents have been studied by Cassell et al. [1994; 1999]. Algorithms and architectures for controlling the non-verbal behaviour of agents, including gaze, are also presented in [Chopra-Khullar and Badler, 1999] and [Novick et al., 1996]. Poggi et al. [2000] provide an interesting basis for implementing eye communication of agents, including gaze, by relating it to the communicative parameters that are involved in a face-to-face interaction, more specifically the conversational actions of the agents and their beliefs, desires and intentions. Most of these authors have focused on getting the appropriate computational models instead of on evaluation.

Before we did our experiment, the work on evaluation of gaze behaviour in mediated communication had been concerned almost only with human-human conversations in videoconferencing and not, to any great extent, with conversations between human and autonomous embodied conversational agents. However, the evaluation work on human-controlled avatars and mediated conversation seemed to provide a promise for reasonable effects in mediated conversations with agents. Hence we were motivated to investigate the effects of implementing different gaze behaviour for Karin, our embodied conversational agent, even despite the fact that in this case the conversation is somewhat unnatural in that users have to type in their parts of the conversation and despite the fact that people may become distracted by the information presented in tables.

Previous work on evaluation in this respect is reported in [Thórisson and Cassell, 1996]. They found that conversations with a gaze informed agent increased ease, believability and efficiency compared to a content-only agent and an agent that produced content and emotional emblems.

Since we did our experiment, some more research on gaze has been published. Fukayama et al. [2002] have implemented a gaze movement model based on observations in the psychological literature. By systematically varying the parameters *amount of gaze*, *mean duration of gaze* and *gaze points while averted*, they have tried to influence the impression management of the agent. They set up an initial experiment to confirm the validity of the gaze movement model and found that the subjects could note the impressions they gained from an eyes-only agent moving its eyes based on their set of gaze parameters.

In our pilot experiment described in the next section, we were not so much interested in the precise rules or the architecture of the system implementing the rules, but rather in the effects on dialogue quality that a simple implementation of the patterns might have. Some of the factors that we looked into are the efficiency of interactions, the way people judge the character of the agents and how they rate the quality of the conversation in general.

3. The Experiment

3.1 Participants, Task and Procedure

We had 48 participants in our experiment. They were all graduate students of the University of Twente, aged between 18 and 25, two thirds were male and one third female. These participants were randomly assigned to one of the three conditions, taking care that the male/female ratio was roughly the same for each.

The participants were given instructions on paper to make reservations for two concerts. During the execution of the task they were left alone in a room monitored by two cameras. After they finished the task they filled out a questionnaire. The questionnaire together with the notes taken when observing the participants through the camera and the time it took for the participants to complete the task were used to evaluate the differences between the three versions of the agent.

3.2 Versions

In the web-accessible version of Karin and the 3D world, visitors have to enter the virtual environment and walk to the reception desk to talk to Karin. In the experiment we started the application so that the participants were positioned face-to-face with Karin immediately. We also left the dialogue box, in which Karin's replies are normally typed out, blank to reduce distraction.

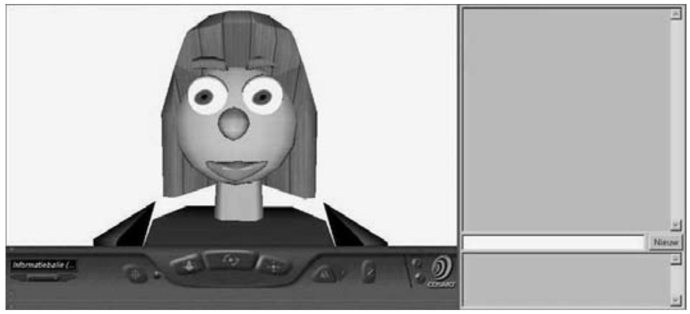

Figure 1.3. Karin as presented in the experiment.

In a face-to-face conversation, people typically look at each other, at objects of mutual interest and blankly into space or at irrelevant objects [Argyle and Cook, 1976]. Karin in our experiment is capable of the following behaviours.

1 Gaze at the visitor (Gaze)

2 Look away from the visitor (Avert)

3 Look at the table of performances (Direct)

The third behaviour refers to a table of performances that can appear below Karin (see Figure 1.2) when performances were retrieved in response of a user query. In this case Karin turns her eyes down to draw attention toward it. This occurs together with Karin saying: “Take a look at the table for a list of the performances.” or something similar.

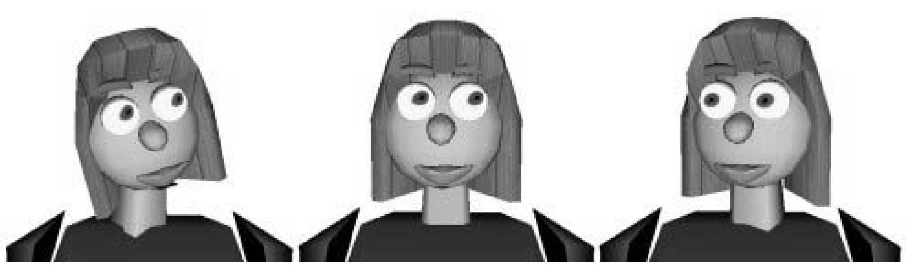

Figure 1.4. Various ways of looking away.

When Karin turns her eyes away from the interlocutor, she will mostly turn her eyes upwards[1] as this is the most typical way of indicating a thinking mode [Poggi et al., 2000]. Several stylistic variations of looking away from the

[1] People typically have a bias towards one direction when averting their eyes in thinking mode. Karin will mostly turn her eyes leftwards.

visitor were implemented. They are accompanied by tilting and rotating the head as is typical for the way humans look away as the pictures above show.

In Table 1.2 a part of a typical conversation is given with indications of how Karin turns her eyes away and towards the human participant. We show the optimal and suboptimal version.

Table 1.2. Sample Dialogue.

		Optimal	Suboptimal
K	Hello, I'm Karin.	Avert	Gaze
	What can I help you with?	Gaze	Gaze
S	Hi. When is the next concert of X?		
K	Just a moment, while I look it up. There are 27 concerts.	Avert	Gaze
	Take a look at the table	Direct	Direct
	For the dates.	Gaze	Gaze
S	I want to book tickets for the concert on November 7.		
K	You want to make a reservation for the Lunch series.	Avert	Gaze
	I have the following information for this series:	Gaze	Gaze
	20 guilders normal rate.	Avert	Gaze
	How many tickets do you want?	Gaze	Gaze

In the *optimal* version Karen averts her eyes at the beginning of a turn for a short period and then starts gazing again. In general Karin's replies are quite short. But some consist of somewhat longer sequences. For instance, when she repeats the information she has so far and also adds a question to initiate the next step in the reservation. This is illustrated by the last reply. In that case, Karin averts her eyes from the speaker to indicate that she is not ready yet and does not want the user to take the turn. We have tried to time eye-movements and information-structure in accordance with the rules described by Cassell et al. [1999]:

- **for each** proposition in the list of propositions to be realized sequentially by a language generator
  - **if** current proposition is thematic
    - **if** beginning of turn or distribution(*.70*) attach a look-away from the hearer
    - **endif**
  - **else if** current proposition is rhematic
    - **if** end of turn or distribution(*.73*) attach a look-toward the hearer

  * **endif**

- **endfor**

In this pseudo-code *distribution(x)* is a randomized function that returns *true* with probability *x*. We did not use this probability function in our algorithm. The length of the utterances was used to determine whether a gaze or avert action could take place. Because the dialogue grammar is not very extensive, we were able to mark it up so we could produce the results as one would expect from the algorithm.

We introduced a second version in which Karin will only stop looking at the user when she directs the users with her eyes to the table with the performances. We will refer to this as the *suboptimal* version. Eye-movements are severely limited in this version. In such a version, there are thus no cues given by the eyes with respect to turn-structure.

In the third version a *random* eye-movement action was chosen at each position at which a specific eye-movement change could occur in the optimal version. This means that there will be more eye-movements than in the suboptimal version, but the movements will not, in general, be related to the conventions described above.

The pictures in Table 1.3 show some of the interaction between Karin (the optimal version) and one participant of the experiment.

In the first shot and second shot we see Karin introducing herself. She gradually tilts her head to the left and turns her eyes away from the user. She immediately turns her head back and resumes eye-contact when she starts her second utterance: the question "What can I help you with?". As soon as this starts the participant puts his fingers on the keyboard, waiting for Karin to end her turn, ready to start typing in his question, as can be seen in the third screen shot. Next the user types in his question, asking whether he can make a reservation. During a brief period, while Karin is asking about details, the user reads the instructions for details about the task he was given. The dialogue manager that takes care of the agent's dialogue intelligence will ask a series of questions to get all the information needed to make a reservation. The last shot shows how Karin looks at the table that lists the performances that match the user's query.

3.3 Measures

In general, we wanted to find out whether participants talking to the optimal version of Karin were more satisfied with the conversation than the other participants. We distinguished between several factors that could be judged: *ease of use, satisfaction, involvement, efficiency, personality/character, naturalness* of eye and head movements and *mental load*. Most of the measures

Table 1.3. Screenshots from an interaction.

K: Hi I am

Karin. What

can I help you with?

P: I wanted to reserve tickets

K: What performance did you want to make a reservation for?

K: For the performances, I refer you to the table.

were judgements on a five point Likert scale (<agree>/<disagree>). A selection of the questions asked is presented in Table 1.4.

Table 1.4. Sample Questions.

Satisfaction
I liked talking to Karin
It takes Karin too long to respond
The conversation had a clear structure
I like ordering tickets this ways
Ease of Use
It is easy to get the right information
It was clear what I had to ask/say
It took a lot of trouble to order tickets
Involvement
I think I looked at Karin about as often as I look to interlocutors in normal conversations
Karin keeps her distance
It was always clear when Karin finished speaking
Personality
I trust Karin
Karin is a friendly person
Karin is quite bad tempered

Some factors were evaluated by taking other measures into account. The time it took to complete the tasks was used, for instance, to measure efficiency. We also asked participants some questions about the things said in the dialogue to judge differences in the attention they had paid to the task.

We were not sure whether participants would be influenced a lot by the differences in the gaze behaviour. However, if there were any effects, we assumed that the optimal version would be most efficient, in that it signals turn-taking mimicking human patterns.

3.4 Results

Efficiency, measured in terms of the time to complete the tasks, was analyzed using a one-way ANOVA test. The table shows the time in minutes. A significant difference was found between the three groups ($F(2,45)=3.80$, $p<.05$). For means and corresponding standard deviations see Table 1.5. To find out which version was most efficient, the groups were compared two by two using t-tests (instead of post-hoc analysis). The optimal version was found to be significantly more efficient than the suboptimal version ($t(30)=-2.31$, $p<.05$, 1-tailed) and the random version ($t(30)=-2.64$, $p<.01$). No significant difference (at 5% level) was found between the suboptimal and the random version.

The main effect of the experimental conditions on the other factors was analyzed using the Kruskal-Wallis test. Answers to questions were recoded such that for all factors the best possible score was 1 and the worst score was

5. The results are summarized in Table 1.5. The table shows significant differences between the versions for ease of use, satisfaction and naturalness of head movement and a marginally significant difference for personality.

Table 1.5. The main effects of experimental condition: means and standard deviations (in parentheses) of the factor scores and the results of the Kruskal-Wallis test.

Factors	Opti	Sub	Ran	χ^2
Ease of use	2.55 (1.31)	3.05 (1.30)	2.66 (1.17)	12.09*
Satisfaction	2.33 (1.20)	2.74 (1.29)	2.79 (1.20)	9.63*
Involvement	3.08 (1.35)	3.47 (1.28)	3.47 (1.17)	3.53
Personality	2.46 (1.21)	2.79 (1.27)	2.79 (1.14)	5.62†
Natural head movement	1.31 (.62)	1.31 (.55)	1.63 (.61)	11.66*
Natural eye movement	1.13 (.39)	1.13 (.49)	1.29 (.58)	3.34
Attention	2.54 (1.27)	3.02 (1.31)	2.63 (1.20)	3.93
Efficiency	6.88 (2.00)	8.88 (2.83)	9.56 (3.56)	-

† p<.10
* p<.01

Two by two comparisons using Mann-Whitney tests pointed out that on the factor *ease of use* the optimal version was significantly better than the suboptimal version (U=6345, p<.001). Users of the optimal version were more *satisfied* than users of the suboptimal and the random version (resp. U=5140, p<.05 and U=4913.5, p<.01). On the factor *personality* the optimal version was better than the random version (U=5261.5, p<.05) and marginally better than the suboptimal version (U=5356.5, p<.10). Both the optimal and the suboptimal agent *moved* their *head* more naturally than the random agent (resp. U=805.5, p<.01 and U=823.5, p<.01). The *eye movements* were found to be marginally better in the optimal version than in the random version (U=1006, p<.10). On the factor *attention* the difference between the optimal version and the suboptimal version was marginally significant (U=910, p<.10). The other comparisons yielded no significant differences.

4. Discussion

The table clearly shows that the optimal version performs best overall. We can thus conclude that even a crude implementation of gaze patterns in turn-taking situations has significant effects. Not only do participants like the optimal version best (satisfaction), find it more easy to use and looking more natural, they also perform the tasks much faster. The more natural version is preferred above a version in which the eyes are fixed almost constantly and a version in which the eyes may move as much as in the optimal situation but do not follow the conventional patterns of gaze.

To measure satisfaction participants were asked to rate how well they liked Karin and how they felt the conversation went in general besides some other questions that relate directly or indirectly to what can be called satisfaction. The participants of the optimal version were not only more satisfied with their version, but they also related more to Karin than the participants of the other versions did as they found her to be more friendly, helpful, trustworthy, and less distant. The differences between the optimal and the suboptimal version seem to correspond to patterns observed in human-human interaction. In the suboptimal version, Karin looks at the visitor almost constantly. Although in general it is the case that people who look more tend to be perceived more favourably, as mentioned above [Kleinke, 1987], in this case the suboptimal version, in which Karin looks at the participants the most of all the versions, is not the preferred one. This, however, is in line with a conclusion of Argyle et al. [1974] who point out that continuous gaze can result in negative evaluation of a conversation partner. This is probably the major explanation why Karin is perceived less favourably as a person in the suboptimal version (as compared to the optimal version). Note that Karin still looks at participants quite a lot in the optimal version as she only looks away at beginning of turns and at potential turn-taking positions when she wants to keep the turn, otherwise she will look at the listener while speaking. She also looks towards the interlocutor while listening. She therefore seems to have found an adequate equilibrium in gazing a lot to be liked but not too much.

When participants have to evaluate how natural the faces behave it appears that the random version scored lower than the other versions but no differences could be noted between the optimal and suboptimal version. Making “the right” head and eye movements or almost no movements are both conceived of as being equally natural, whereas random movements are judged less natural. What is interesting, however, is that these explicit judgements on the life-likeness of the behaviour of the agents do not reflect directly judgments on other factors. The random version may be rated as less natural than the others but in general it does not perform worse than the suboptimal version. For the factor ease of use it is judged even significantly better than the suboptimal version. Does this mean that having regular movements of the eyes instead of almost fixed eyes is the important cue here? On the other hand, the difference in this rating (which is gotten from judgments on a question like “was it easy to order tickets”) is not in line with the real amount of time people actually spent on the task. Though the random version is judged easy to use, it takes the participants using it the most time to complete the tasks.

The optimal version is clearly the most efficient in actual use. This gain in efficiency might be a result of the transparency of turn-taking signals; i.e. the flow of conversation may have improved as one would assume when regulators like gaze work appropriately. But the gain might also have been a result, indirectly, of the increased involvement in the conversation of the participants that used the optimal version. Whatever is cause or effect is difficult to say. We have an indication that the different gaze patterns had some impact not just on overall efficiency but also on the awareness of participants about when Karin was finishing her turn. We have some rough figures on the number of times participants started their turn before Karin was finished with hers. In almost all of these cases this slowed down the task, because participants would have to redo change their utterance midway.

Table 1.6. Number of participants interrupting Karin.

	Opt	Sub	Ran
Often/Regularly		5	4
Sometimes	4	2	3
Never	12	9	9

These figures are not conclusive, but give an indication that at least in the optimal version, participants did seem to take into account the gaze behaviour of Karin as part of the cues that regulate turn-taking behaviour.

5. Conclusion

In face-to-face conversations between human interlocutors, gaze is an important factor in signalling interpersonal attitudes and personality. Gaze and mutual gaze also function as indicators that help in guiding turn-switching. In the experiment that we have conducted, we were interested in the effects of implementing a simple strategy to control eye-movements of an artificial agent at turn-taking boundaries.

The crude rules that we have used are sufficient to establish significant improvements in communication between humans and embodied conversational agents. So, therefore, the effort to investigate and implement human-like behaviour in artificial agents seems to be well worth the investment.

References

Argyle, M. (1993). *Bodily Communication*. Routledge, second edition.

Argyle, M. and Cook, M. (1976). *Gaze and Mutual Gaze*. Cambridge: Cambridge University Press.

Argyle, M. and Dean, J. (1972). Eye Contact, Distance and Affiliation. In Laver, J. and Hutcheson, S., editors, *Communication in Face to Face Interaction*, pages 155–171. Penguin.

Argyle, M., Lefebre, L., and Cook, M. (1974). The Meaning of Five Patterns of Gaze. *European Journal of Social Psychology*, 4(2):125–136.

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., and Stone, M. (1994). Animated Conversation. Rule Based Generation of Facial Expression, Gesture and Spoken Intonation for Multiple Conversational Agents. *Computer Graphics*, pages 413–420.

Cassell, J., Sullivan, J., Prevost, S., and Churchill, E., editors (2000). *Conversational Agents*. MIT Press.

Cassell, J., Torres, O., and Prevost, S. (1999). Turn Taking vs. Discourse Structure. In Wilks, Y., editor, *Machine Conversations*, pages 143–154. Kluwer.

Chopra-Khullar, S. and Badler, N. I. (1999). Where to Look? Automating Attending Behaviors of Virtual Human Characters. In *Proceedings of Autonomous Agents*, pages 9–23, Seattle.

Colburn, R. A., Cohen, M. F., and Drucker, S. M. (2000). Avatar Mediated Conversational Interfaces. Technical Report MSR-TR-2000-81, Microsoft.

Fukayama, A., Ohno, T., Mukawa, N., Sawaki, M., and Hagita, N. (2002). Messages Embedded in Gaze of Interface Agents. Impression Management with Agent's Gaze. In *Proceedings of Conference on Human Factors in Computing Systems (CHI)*, pages 41–48, Minneapolis, Minnesota, USA. ACM.

Garau, M., Slater, M., Bee, S., and Sasse, M.A. (2001). The Impact of Eye Gaze on Communication Using Humanoid Avatars. In *Proceedings of Conference on Human Factors in Computing Systems (CHI)*, pages 309–316, Seattle, Washington, USA.

Heylen, D., Nijholt, A., and Poel, M. (2001). Embodied Agents in Virtual Environments: The Aveiro Project. In *Proceedings of European Symposium on Intelligent Technologies, Hybrid Systems and their Implementation on Smart Adaptive Systems*, pages 110–111, Tenerife, Spain. Verlag Mainz, Wissenschaftsverlag Aachen.

Kendon, A. (1990). Some Functions of Gaze Direction in Two-Person Conversation. In *Conducting Interaction*, pages 51–89. Cambridge: Cambridge University Press.

Kleinke, C. L. (1987). Gaze and Eye Contact: A Research Review. *Psychological Bulletin*, 100:78–100.

Nijholt, A. and Hulstijn, J. (2000). Multimodal Interactions with Agents in Virtual Worlds. In Kasabov, N., editor, *Future Directions for Intelligent Information Systems and Information Science*, pages 148–173. Physica-Verlag.

Novick, D. G., Hansen, B., and Ward, K. (1996). Coordinating Turn-Taking with Gaze. In *Proceedings of International Conference on Spoken Language Processing (ICSLP)*, pages 1888–1891, Philadelphia, Pennsylvania, USA.

Poggi, I., Pelachaud, C., and de Rosis, F. (2000). Eye Communication in a Conversational 3D synthetic Agent. Special Issue on Behavior Planning for Life-Like Characters and Avatars. *AI Communications*, 13(3):169–181.

Thórisson, K. R. and Cassell, J. (1996). Why Put an Agent in a Body: The Importance of Communicative Feedback in Human-Humanoid Dialogue. In *Proceedings of Lifelike Computer Characters*, pages 44–45, Utah, USA.

Vertegaal, R. (1999). The GAZE Groupware System: Mediating Joint Attention in Multiparty Communication and Collaboration. In *Proceedings of Conference on Human Factors in Computing Systems (CHI)*, pages 294–301, Pittsburgh, Pennsylvania, USA. ACM.

Vertegaal, R., Slagter, R., van der Veer, G., and Nijholt, A. (2001). Eye Gaze Patterns in Conversation. There is More to Conversational Agents than Meets the Eyes. In *Proceedings of Conference on Human Factors in Computing Systems (CHI)*, pages 301–308, Seattle, Washington, USA.