Experimenting with the Gaze of a Conversational Agent

Dirk HEYLEN **Ivo VAN ES** **Anton NIJHOLT** **Betsy VAN DIJK**
Computer Science, University of Twente
PoBox 217
7500 AE Enschede, The Netherlands,
{heylen,es,anijholt,bvdijk}@cs.utwente.nl

Abstract

We have carried out a pilot experiment to investigate the effects of different eye gaze behaviors of a cartoon-like talking face on the quality of human-agent dialogues[1]. We compared a version of the talking face that roughly implements some patterns of human-like behavior with two other versions. We called this the optimal version. 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 optimal agent appreciated the agent more than participants that conversed with the other agents. Conversations with the optimal version proceeded more efficiently. Participants needed less time to complete their task.

Introduction

Research on embodied conversational agents is carried out in order to improve models and implementations simulating aspects of human-like conversational behavior as best as possible. Ultimately, one would like 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-channeling, 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 like tutoring (Cassell, 2001).

In our research laboratory we started to develop spoken dialogue systems some years ago. We focused on 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 systems. 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. 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

[1] Short 2 page papers related to this experiment were submitted to the CHI 2002 conference (Minneapolis) and AVI (Trento) and accepted for presentation. We have benefitted greatly from comments made by anonymous reviewers to these versions.

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 behavior including facial expressions, posture and gesture, and gaze (which is the topic of this paper).
In the next section of this paper we will discuss some aspects of the function of gaze in face-to-face conversations between humans and in mediated forms. Next we describe our experiment and discuss the outcome.

1. Functions of (mutual) gaze

The function of gaze in human-human, face-to-face dialogues has been studied quite extensively (see Argyle and Cook (1976) and many other publications mentioned in the references). 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 behavior 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 have been considered.
Gaze has been shown to serve a number of functions in human-human interaction (Kendon, 1990). It helps to regulate the flow of conversation and plays an important role in ensuring smooth turn-taking behavior. 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. Duration and types of gaze communicate the nature of the relationship between the interlocutors.
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 forced 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 and on other forms of mediated 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 behavior. Evaluations of the effects of gaze on the quality of interactions in mediated conversation (mostly 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 gaze behavior of agents or avatars in human-agent or human-avatar communication has noticeable effects on the way communication proceeds. This made us curious about our own situation with the agent Karin. We wondered whether implementing some kind of human-like rules for gaze behavior would have any effects given her somewhat limited dialogue functionality, her cartoon-like face, the somewhat unnatural way of input that lets users type in their questions only instead of using speech and the fact that the face is only one modality amongst others that is used to present information. We therefore set up our experiment which is further described in Section 3.

1.1 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.
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.

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

Among the common subjective interpretations of eye contact have been found friendship, sexual

attraction, hate and a struggle for dominance. Gaze levels are also higher in those who are extroverted, dominant or assertive, and socially skilled. 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 behavior and the relation to (mutual) gaze have been the subject of several investigations. In our experiment we wanted to focus the attention on the way appropriate rules of gazing of the agent would improve the quality of the conversation. However, we also wanted to see whether the different patterns that we had chosen would affect the way our agent was liked or disliked.

Studying the patterns in gaze and turn-taking behavior, 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/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. the table above). 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.

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 paper, 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), than the speaker will always look towards the hearer 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.

1.2 Mediated Conversation

Several researchers have investigated the effects of implementing gaze behavior in conversational agents or in other forms of mediated conversation. In videoconfering 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 appear to look at each other on the screen (see also Vertegaal et al., 2001).

Garau et al. (2001) describe an experiment with dyadic conversation between humans in 4 mediated conditions: video, audio-only, random-gaze avatars and informed gaze avatars (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 behavior of avatars during a conversation. It appears from this and similar experiments that participants while not consciously aware of the differences in the avatar’s gaze behavior will still react differently (subliminally).

In the context of embodied conversational agents, rules for gaze behavior of agents have

been studied by Cassell et al. (1994, 1999). Algorithms and architectures for controlling the non-verbal behavior, including gaze, of agents are also presented in Chopra et al. (2001) and Novick et al. (1996). These have focussed mainly on getting the appropriate computational models instead of on evaluation. 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.

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 wanted to look 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.

Although the work on evaluation of gaze behavior has not been concerned to any great extent with autonomous embodied conversational agents, the evaluation work on human-controlled avatars and mediated conversation seemed to provide a promise for reasonable effects in mediated conversations with agents in general and even with our agent Karin whom users have to interact with by typing in their utterances and who presents information also in the form of tables.

2 Our experiment

In our experiment we compared three versions of Karin that differed with respect to gaze behavior. We had 48 participants each carry out two ticket reservation tasks with one version of Karin. After they had finished, they filled out a questionnaire. Together with some other measures (such as the time it took them to complete the tasks) this data was used to evaluate the implementations on a number of factors.

2.1 Participants, task and procedure

The 48 participants in our experiment 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.

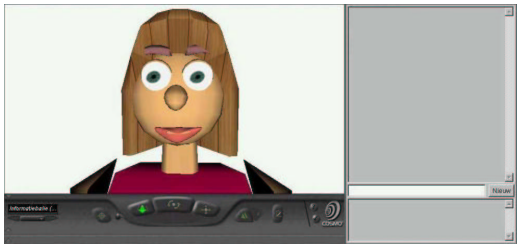
The participants were given the task to make reservations for two concerts. During the execution of the task they were left alone in a room monitored by two cameras. After they finished the task they filled out a questionnaire. The questionnaire together with the notes taken when observing the particpants 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.

2.2 Versions

In the following table 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. In the “direct” behavior, Karin turns her eyes downwards, towards a table that contains information about the performances.

		Optimal	Sub
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 will avert her eyes at the beginning of a turn for a short period and then start gazing again. In general Karin's replies are quite short. But some consist of some 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 will avert 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).



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. Eye-movements are therefore limited severely in this version. 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.

2.3 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*, *effiency*, *personality/character*, *naturalness* (of eye and head movements) and *mental load*. Most of the measures were judgements on a five point Likert scale (<agree>/<disagree>). A selection of the questions asked is presented below. 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 asked participants some questions about the things said in the dialogue to judge differences in attention (mental load).

Satisfaction

I <liked> / <didn't like> 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

We were not sure whether participants would be influenced a lot by the differences in the gaze behavior. 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.

2.4 Results

Efficiency was analyzed using a one-way ANOVA test. A significant difference was found between the three groups ($F(2,45)=3.80$, $p<.05$). For means and corresponding standard deviations see the table below. 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 subobtimal 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 worse score was 5. The results are summarized in the table. 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.

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	X^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
Mental load	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)	-

† $p<.10$ * $p<.05$ ** $p<.01$

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 *mental load* the difference between the optimal version and the suboptimal version was marginally significant (U=910, $p<.10$). The other comparisons yielded no significant differences.

3 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, they also perform the tasks much faster and tend to be more involved in the conversation. 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 behind the negative effect on how Karin is perceived as a person in this 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 equilibrum 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 behavior 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 questions like "does it take Karin long to respond", "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 transparancy 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.

	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 paid more attention to Karin than in the other versions.

4 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 effect significant improvements in communication between humans and embodied conversational agents. So, therefore, the effort to investigate and implement human-like behavior in artificial agents seems to be well worth the investment.

References

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

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

M. Argyle, J. Dean (1972) Eye contact, distance and affiliation. Reprinted in. J. Laver, S. Hutcheson (eds.) *Communication in Face to Face Interaction*, Penguin [1962 original] (p. 155-171).

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

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

J. Cassell, O. Torres, S. Prevost (1999). Turn Taking vs. Discourse Structure, in *Machine Conversations* (p. 143-154).

J. Cassell, J. Sullivan, S. Prevost, E. Churchill (eds.) (2000) Embodied Conversational Agents, MIT Press.

S. Chopra-Khullar, N.I. Badler (1999). Where to look? Automating attending behaviors of virtual human characters. In: *Proceedings of Autonomous Agents*. Seattle.

R.A. Colburn, M.F. Cohen, S.M. Drucker (2000) Avatar Mediated conversational interfaces. Microsoft Technical Report. MSR-TR-2000-81. July 2000.

M. Garau, M. Slater, S. Bee, M.A. Sasse (2001) The impact of eye gaze on communication using humanoid avatars. In: *CHI* 2001 (p. 309-316).

D. Heylen, A. Nijholt & M. Poel (2001) Embodied agents in virtual environments: The Aveiro project. In: Proceedings European Symposium on

Intelligent Technologies, Hybrid Systems and their implementation on Smart Adaptive Systems, Tenerife, Spain, December 2001, Verlag Mainz, Wissenschaftsverlag Aachen, 110-111.

A. Kendon (1990) Some functions of gaze direction in two-person conversation. Reprinted in: *Conducting Interaction*, Cambridge University Press, Cambridge (p. 51-89).

C.L. Kleinke (1987). Gaze and Eye Contact: a research review. In: *Psychological Bulletin*, 100 (p. 78-100).

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

D.G. Novick, B. Hansen, K. Ward (1996) Coordinating Turn-Taking with Gaze. In: *Proceedings ICSLP*.

K.R. Thórisson, J. Cassell (1996) Why Put an Agent in a Body: the importance of communicative feedback in human-humanoid dialogue. Presented at Lifelike Computer Characters, Utah, October 1996.

R. Vertegaal (1999) The GAZE Groupware system: Mediating Joint Attention in Multiparty Communication and Collaboration. In: *Proceedings of CHI*'99, Pittsburgh, ACM Press (p. 294-301).

R. Vertegaal, R. Slagter, G. van der Veer, A. Nijholt (2001) Eye Gaze Patterns in Conversation. There is more to conversational agents than meets the eyes. In: *Proceedings of CHI 2001 Anyone. Anywhere*. ACM.