Gaze Behavior of Talking Faces Makes a Difference

Ivo van Es, Dirk Heylen, Betsy van Dijk, Anton Nijholt
Department of Computer Science
University of Twente, the Netherlands
PoBox 217
7500 AE Enschede
{es,heylen,bvdijk,anijholt}@cs.utwente.nl

ABSTRACT
We present the results of an experiment investigating the effects of a talking head's gaze behavior on the user's quality assessment of the interface. We compared a version that used life-like rules for gazing with a version that would keep its eyes fixed on the visitor most of the time, and a random version. We found significant differences between these gaze algorithms in terms of ease of use, efficiency and other quality factors.

Keywords
Conversational agents, gaze, evaluation.

INTRODUCTION
One of the non-verbal communication systems that has been looked into by a number of researchers on embodied conversational agents is *gaze* [1,4,5,6,8,9,10]. In human-human, face-to-face conversations, typical patterns can be observed in the way interlocutors make eye contact or look away. Gazing at the other or averting gaze can be used consciously to signal information or it can involuntarily provide cues about interpersonal relations such as liking or dominance, and personality characteristics like shyness. By looking away from the speaker a hearer might show a lack of interest. These effects have been described extensively in the literature on non-verbal communication ([1]).

In our research we were interested in the effects of simulating the correlation between patterns of gaze behavior, turn structure and information structure (9]). We therefore focused on gaze patterns at turn-boundaries. In general, when starting to speak, a speaker will often avert the eyes from the listener (to concentrate on what he is going to say). At the end of the turn, the speaker will typically direct gaze to the listener again, in order to signal the end of the turn and to provide the hearer with the opportunity to take the turn. This is the basic pattern that we wanted to investigate. We also took into account the information structure (theme/rheme) of the sentences uttered by the agent. The main question was whether conversations with our embodied conversational agent would improve qualitatively if the agent followed this pattern (see also [8]).

CHI 2002, April 20-25, 2002, Minneapolis, Minnesota, USA.
ACM 1-58113-454-1/02/0004.

EXPERIMENT
In our experiment we compared three versions of our agent Karin that differed in gaze behavior. Karin has a human-like appearance that has been realized using VRML. She allows a simple, but nevertheless mixed-initiative natural language dialogue with a user, during which information can be obtained about theatre performances and reservations can be made [7]. We had 48 subjects each carry out two reservation tasks with one version of Karin. After they had finished, they filled out a questionnaire.

Versions
In the so-called "optimal" version, Karin turns her eyes away from the visitor when she starts to speak and looks at the speaker just before ending her turn. In the second "suboptimal" version, Karin keeps her eyes fixed on the visitor most of the time. In the third version Karin chooses a gaze action (look towards, look away, direct eyes) on key positions (beginning of turn, end of turn) at random.

PARTICIPANTS
Subjects were all graduate students, aged between 19 and 22, two thirds were male, one third female. There were 48 testsubjects (16 per version), randomly distributed over the 3 versions, taking care that the ratio male/female testsubjects for each version was roughly the same.

Task
Testsubjects were given a letter in which the proceedings were described. They were given the task to make reservations for 2 concerts. They used the same version of Karin for both reservations.

Factors – Questions – 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 subjects. We distinguished between several factors that could be judged: *ease of use*, *satisfaction*, *involvement*, *efficiency*, *personality*, *naturalness* (of eye and head movements). Most of the measures were judgements on a five point Likert scale. 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.

Results

The results of the questionaires and other measures were analysed statistically (using the Kruskal Wallis, Mann-Whitney and the t-test). The main results can be summarised as follows, where O=optimal, S=suboptimal, R=random, >S-R=significantly (5% significance level) better than S and R, ≥ means quite different but not significant (at 10% significance level).

	O	S	R
Satisfaction	> S-R		
Ease of Use	> S		> S
Involvement	≥ S-R		
Character/Personality	> S-R		
Natural head-movement	> R	> R	
Natural eye-movement	≥R	≥ R	
Efficiency	>R ≥S	≥R	

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 subjects 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, subjects 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 subjects of the optimal version were not only more satisfied with their version, but they also related more to Karin than the testsubjects 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. In [2] it is pointed 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.

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 noticed 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 slightly 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, though this subjective rating is not reflected in the efficiency measure. It appears that eye-movements may not be registered consciously but still have effects subliminally as was also noticed in [5]. 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 if a regulator like gaze works appropriately. We have some rough but inconclusive figures on the number of times subjects started their turn before Karin was finished with hers, corrobarating this.

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 found significant differences in the effects of implementing different strategies to control eye-movements of an artificial agent at turn-taking boundaries. The most "life-like" version scored best.

REFERENCES

1. M. Argyle, M. Cook. *Gaze and Mutual Gaze*. Cambridge University Press, Cambridge, 1976.
2. M. Argyle, L. Lefebre, M. Cook. The Meaning of Five Patterns of Gaze. In: *European Journal of Social Psychology*, 4(2), 1974 (p.125-136).
3. J. Cassell, C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, M. Stone. Animated Conversation. In: *Computer Graphics* 1994 (p. 413-420).
4. S. Chopra-Khullar, N.I. Badler. Where to look? Automating attending behaviors of virtual human characters. In: *Proc. Autonomous Agents*. Seattle, 1999.
5. R.A. Colburn, M.F. Cohen, S.M. Drucker. Avatar Mediated conversational interfaces. Microsoft Technical Report. MSR-TR-2000-81. July 2000.
6. M. Garau, M. Slater, S. Bee, M.A. Sasse. The impact of eye gaze on communication using humanoid avatars. In: *CHI* 2001 (p. 309-316).
7. A. Nijholt, J. Hulstijn. Multimodal Interactions with Agents in Virtual Worlds. In: N. Kasabov (ed.) *Future Directions for Intelligent Information Systems and Information Science*, Physica, 2000 (p. 148-173).
8. K.R. Thórisson, J. Cassell. Why Put an Agent in a Body. Lifelike Computer Characters, Utah, 1996.
9. O. Torres, J. Cassell, S. Prevost. Modeling gaze behavior as a function of discourse structure. First International Workshop on Human Computer Conversations. Bellagio, Italy, 1997.
10. R. Vertegaal, R. Slagter, G. van der Veer, A. Nijholt. Eye Gaze Patterns in Conversation. In: *Proceedings of CHI 2001 Anyone. Anywhere*. ACM