Making agents gaze naturally - Does it work?

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

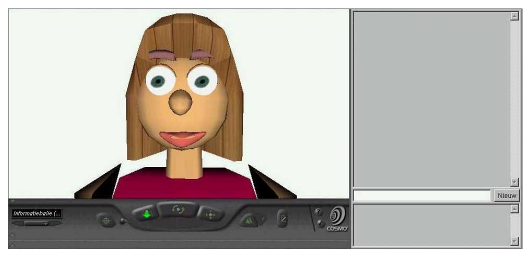
ABSTRACT
We investigated the effects of varying eye gaze behavior of an embodied conversational agent on the quality of human-agent dialogues. In an experiment we compared three versions of an agent: one with gaze behavior that is typically found to occur in human-human dialogues, one with gaze that is fixed most of the time, and a third version with random gaze behavior. The versions were found to yield significant differences in efficiency of the dialogues and in user satisfaction, amongst others.

Keywords
Conversational agents, gaze, non-verbal communication.

1. INTRODUCTION
One of the aims in building embodied conversational agents is to emulate human behavior in the agents. The general aim of our research is to qualitatively improve the communication between the virtual agent and a human, by simulating the way humans look at each other during a conversation. *Gaze* is an important non-verbal communication factor that has been subject to several investigations ([1,3,4,5,6,7,8,9]). In human-human, face-to-face conversations, typical patterns have been observed in the way interlocutors make eye contact or look away ([2]). The literature on non-verbal communication describes several functions of gaze behavior extensively ([1]). Gazing at the other or averting gaze can be used consciously to signal information. For instance, by looking away from the speaker a hearer might show a lack of interest. But gaze behavior can also subconsciously provide cues about interpersonal relations such as liking or dominance, and personality characteristics like shyness.

Patterns of gaze behavior, turn structure and information structure are correlated ([8]). In our study 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 give the hearer the opportunity to take the turn. This is the basic pattern that we simulated in the so-called 'optimal' version of the virtual agent. We also took into account gaze behavior that is known to occur at the thematic and rhematic parts of utterances. The thematic part links the utterance with previously uttered or contextualised information, whereas the rhematic part offers new information to the listener.

2. EXPERIMENT
The purpose of the experiment was to examine the impact of the gaze behavior of a virtual agent on the human-agent dialogue. For the experiment we used the embodied conversational agent Karin, a receptionist in the Virtual Music Center (VMC). Because of the web-based design, Karin has a cartoon face and visitors can ask information and buy tickets by typing natural language utterances. Karin uses text-to-speech synthesis to read out her answers (see http://parlevink.cs.utwente.nl/).



Independent variables In the experiment we compared three versions of Karin that differed in gaze behavior. In the so-called 'optimal' condition the global idea of the gaze behavior is that 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 'suboptimal' condition, Karin keeps her eyes fixed on the visitor most of the time. In the 'random' condition Karin randomly chooses between looking towards and looking away from the visitor.

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 condition.

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 from the camera view and the time it took for the subjects to complete the task were used to evaluate the differences between the three versions of the agent.

Measures In general, we wanted to find out whether implementing human-like gaze behavior in the optimal version of Karin improved the quality of the dialogue. We distinguished between several factors that could be judged: *ease of use,* *satisfaction, involvement, efficiency, personality* (amongst others trustworthy, friendly, aggressive or not) *naturalness* (of eye and head movements), and *mental load*. All the measures except efficiency were judgements on a five point Likert scale, each factor measured by several (3-8) questions. The time it took to complete the task was used to measure efficiency.

3. 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 table 1. 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 table 1.

Table 1. 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	Optimal	Sub-optimal	Random	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
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)	-
† $p<.10$	* $p<.05$	** $p<.01$		

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. 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.

We can thus conclude that even a crude implementation of gaze patterns in a web-based agent in turn-taking situations has significant effects. Not only do subjects like the optimal version best and find it easier to use and more natural, they also perform the tasks faster. On most factors the optimal version is preferred above the version in which the eyes are fixed almost constantly and above the random version, in which the head and eyes may move as much as in the optimal situation but do not follow the conventional patterns of gaze.

4. CONCLUSIONS

In face-to-face conversations between human interlocutors, gaze is an important factor in signaling 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 human-like gaze behavior in embodied conversational agents. Significant differences were found between the most “life-like” version and the other two versions. So, the effort to implement human-like behavior may be well worth the investment.

5. REFERENCES

[1] M. Argyle, M. Cook. *Gaze and Mutual Gaze*, Cambridge University Press, Cambridge, 1976.

[2] J. Cassell, O. Torres, S. Prevost. Turn Taking vs. Discourse Structure: how best to model multimodal conversation, in *Machine Conversations*, 1999 (p. 143-154).

[3] S. Chopra-Khullar, N.I. Badler. Where to look? Automating attending behaviors of virtual human characters, in *Proc. Autonomous Agents*, Seattle, 1999 (p. 16-23).

[4] R.A. Colburn, M.F. Cohen, S.M. Drucker. Avatar Mediated conversational interfaces. Microsoft Technical Report, SR-TR-2000-81, July 2000.

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

[6] D.G. Novick, B. Hansen, K. Ward. Coordinating Turn-Taking with Gaze, in *Proceedings ICSLP*, 1996.

[7] K.R. Thórisson, J. Cassell. Why Put an Agent in a Body. Lifelike Computer Characters, Utah, 1996.

[8] 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.

[9] R. Vertegaal, R. Slagter, G. van der Veer, A. Nijholt. Eye Gaze Patterns in Conversation, in *Proceedings of CHI 2001 Anyone. Anywhere*. ACM, 2001 (p.301-308).