# Why Conversational Agents Should Catch the Eye

**Roel Vertegaal**
Queen's University
Canada
*roel@acm.org*

**Robert Slagter**
Telematics Institute
The Netherlands
*slagter@telin.nl*

**Gerrit van der Veer**
Vrije Universiteit
The Netherlands
*gerrit@acm.org*

**Anton Nijholt**
Twente University
The Netherlands
*anijholt@cs.utwente.nl*

**ABSTRACT**
We studied whether the gaze direction of users indicates whom they are speaking or listening to in multiparty conversations. Results show when someone is listening or speaking to individuals, there is indeed a high probability that the person looked at is the person listened ($p$=88%) or spoken to ($p$=77%). We implemented these findings in a multi-agent conversational system that uses eye input to gauge which agent the user is listening or speaking to.

**KEYWORDS**: Conversational Agents, Gaze, Eye tracking.

**INTRODUCTION**
The ability to communicate without words plays an important part in our everyday use of language. Not only do conversational cues such as gestures, facial expressions, looks and tone of voice often determine the meaning of the words we use, such nonverbal expressions may also play an essential role in regulating the conversational process [1]. We are now beginning to see how the lack of support for nonverbal conversational cues may limit the usability of speech recognition and production systems [2]. In this paper, we focus on one particular problem: knowing when the system is being addressed or expected to speak. This problem becomes apparent particularly in multi-agent, multi-user environments, such as our Virtual Theatre [5]. The Virtual Theatre is an animated 3D VRML model of a theatre, in which users can see previews of shows and book tickets through conversational agents. Different agents are used for different queries: to ease contextual knowledge requirements for the system, the embodiment of each agent is used as a metaphor for its functionality. However, the ability to speak to multiple agents means users as well as agents should have a means of establishing who is talking to whom. It has long been presumed that gaze directional cues are an important source of such information in human conversation [6]. In order to verify how well looking behavior of users predicts their conversational attention, we tracked where people looked in normal face-to-face group conversations. We will first discuss previous empirical work. We then discuss our experiment, and our work towards a multi-agent conversational system that can observe and use gaze directional cues.

**PREVIOUS WORK**
According to Kendon [3], in two-person (dyadic) conversations, seeking or avoiding to look at the face of the conversational partner (i.e., *gaze*) serves at least four functions: (1) to regulate the flow of conversation; (2) to provide visual feedback; (3) to communicate emotions and relationships; and (4) to improve concentration by restriction of visual input. In the early seventies, Argyle [1] estimated that when two people are talking, about 60 % of conversation involves gaze, and 30 % involves mutual gaze (or eye contact). People look nearly twice as much while listening (75%) as while speaking (41%). In general, gaze seems closely linked with speech. According to Kendon [3], person A tends to look away as she begins speaking, and starts to look more at her interlocutor B as the end of her utterance approaches. This pattern should be explained from two points of view. Firstly, in looking away at the beginning, person A may be withdrawing her attention from person B in order to concentrate on what she is going to say. When she approaches the end of her utterance, the subsequent action will depend largely upon how person B is behaving, necessitating person A to seek information about her interlocutor. Secondly, these changes in gaze function as signals to person B. In looking away at the beginning, person A signals that she is about to begin speaking, forestalling responses from person B. Similarly, in looking at person B towards the end of her utterance, she may signal that she is now ceasing to talk yet still has attention for him, effectively offering the floor to person B. So how do these findings hold in a multiparty situation? In one of the few (unpublished) studies on gaze behavior in groups, Weisbrod [9] found that subjects gazed over 70% of their speaking time, but only 47% of their listening time. Kendon attributed this reversal of the pattern observed in dyadic studies to the need to make clear to whom one is speaking [3]. When addressing a group, the speaker cannot look at all individuals simultaneously. To avoid too large a drop in gaze per individual, the speaker would increase overall gaze. In order to verify these assumptions, we performed an experiment in which we compared time spent gazing at individuals spoken or listened to with time spent gazing at *others* in four-person group conversations. Our first hypothesis was:

***H1****"On average, significantly more time is spent gazing at the individual one listens or speaks to, than at others"*

To make sure Hypothesis 1 would still hold in cases where visual attention is divided, we added a second hypothesis:

***H2****"On average, significantly more time is spent gazing at each person when addressing a group of three, than at others when addressing a single individual"*

**METHODS**
Our experiment applied a within-subjects design in which all variables were measured, rather than controlled. 7 four-person groups discussed current-affairs topics in face-to-face meetings. Subjects participated in four 8-minute sessions: one in which we recorded where they looked using a desk-mounted eyetracker [4], and three in which they were conversational partners only. We also registered speech activity of each discussant. The location of the conversational partners' faces was determined by tracking subject fixations at the eyes of each partner before each

In Summary of ACM CHI 2000 Conference on Human Factors in Computing, The Hague.

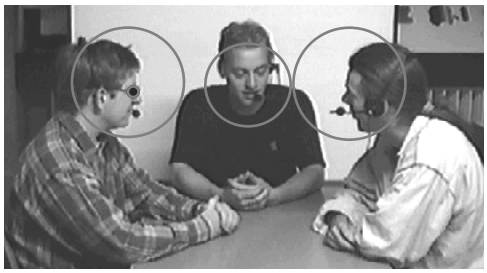

***Figure 1.*** *When subject fixations (the black dot) hit one of the circles, gaze at the corresponding person was registered.*

session. We then fitted the largest possible non-overlapping circles around the mean locations of their eyes (see Figure 1). During the session, subject fixations within a circle would register gaze for that partner. After each session, the tracked subject was asked to watch a video of the discussion to score whom she had spoken or listened to. When the subject thought she had been listening or speaking, she would press one or more keys on a keyboard indicating the person(s) she had listened or spoken to. Subjects were trained for this task using a pre-recorded video, and scores were corrected for response time (see [7]).

## ANALYSIS AND RESULTS

Results were calculated over 24 sessions, with 5 female and 19 male subjects. Only the last 5 minutes of each session were analyzed. We used automated analysis only, verified by human observers. A fuzzy algorithm analyzed the speech patterns of individuals to determine speaker turns (see [7]). By combining this information with the subject's conversational attention scores, we could calculate, for each moment in time, not just whether the subject was listening or speaking, but also whom he was listening or speaking to. For each session and for each conversational partner, we then calculated the mean percentage of time in which the subject fixated his gaze within the facial region of that partner while (a) speaking to that partner, (b) listening to that partner or (c) speaking to all three. Resulting percentages where averaged across partners and subjects, and are presented in Table 1.

| **Variable** | *Listening to individual* | *Addressing individual* | *Addressing all three* |
|---|---|---|---|
| *Gaze at individual* | 62.4 *(3.8)* | 39.7 *(4.7)* | 19.7 *(1.8)* |
| *Gaze at others* | 8.5 *(1.2)* | 11.9 *(2.4)* | |
| *Gaze at all three* | | | 59.0 *(5.4)* |

***Table 1.*** *Means and std. errors for percentages of time spent by subjects gazing at partners in the last 5 session minutes.*

Planned comparisons showed that subjects gazed 7.3 times more at the individual listened to (62.4%), than at others (8.5%) ($t(23)=12.92$, $p<.001$, 1-tailed). They gazed 3.3 times more at an addressed individual (39.7%), than at others (11.9%) ($t(23)=5.2$, $p<.001$, 1-tailed), thus confirming Hypothesis 1. A second planned comparison showed that subjects gazed approximately 1.7 times more at an individual when addressing all three (19.7%), than at others when addressing a single individual (11.9%) ($t(22)=2.71$, $p<.01$, 1-tailed), thus confirming Hypothesis 2.

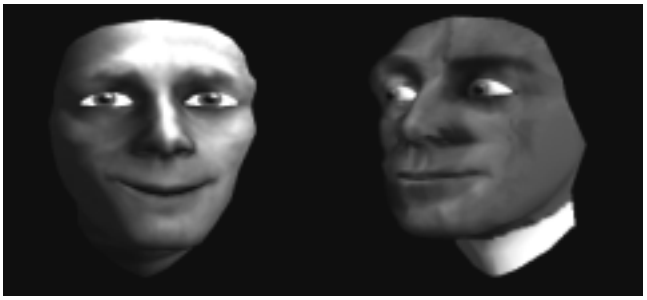

***Figure 2.*** *FRED prototype with two conversational agents.*

## DISCUSSION AND CONCLUSIONS

We can indeed consider gaze behavior to be a very good predictor of people's conversational attention. When someone is listening or speaking to individuals, there is a high probability that the person he looks at is the person he listens (88% chance) or speaks to (77% chance). Although levels of gaze per individual drop when addressing larger groups of three, each person still receives 1.7 times more gaze than could be expected had he not been addressed. This means that the user's eye gaze can form a reliable source of input for conversational systems that need to establish whether the user is speaking or listening to them.

## APPLICATIONS: FRED, AN ATTENTIVE AGENT

We are currently working to implement our findings in FRED, a multi-agent conversational system that establishes where the user looks by means of a desk-mounted LC Technologies eyetracking system [4]. In FRED, multiple conversational agents can be embodied by means of 3D texture-mapped models of humanoid faces. Based on work by Waters [8], muscle models are used for generating accurate 3D facial expressions. The system uses our SCHISMA speech recognition and production engine to converse with the user [5]. Each agent is capable of detecting whether the user is looking at it, and combines this information with speech data to determine when to speak or listen to the user. To help the user regulate conversations, agents generate gaze behavior as well. This is exemplified by Figure 2. In this example, the agent speaking on the left is the focal point of the user's eye fixations. The right agent observes that the user is looking at the speaker, and signals it does not wish to interrupt by looking at the left agent, rather than the user.

## REFERENCES


1. Argyle, M. and Cook, M. *Gaze and Mutual Gaze*. London: Cambridge University Press, 1976.
2. Cassell, J., Bickmore, T., *et al.* Embodiment in Conversational Interfaces: Rea. In *Proceedings of CHI'99.* Pittsburgh: ACM, 1999.
3. Kendon, A. Some Function of Gaze Direction in Social Interaction. *Acta Psychologica* 32, 1967, pp. 1-25.
4. LC Technologies. *http://www.eyegaze.com*, 1997.
5. Nijholt, A., v.d. Hoeven, G., *et al.* SCHISMA: A Natural Language Accessible Theatre Information and Booking System. In *Proceedings of First Workshop on Applications of Natural Language to Databases.* Versailles, France, 1995, pp. 271-285.
6. Vertegaal, R. The GAZE Groupware System: Mediating Joint Attention in Multiparty Communication and Collaboration. In *Proceedings of CHI'99.* Pittsburgh: ACM, 1999.
7. Vertegaal, R. *Look Who's Talking to Whom*. PhD Thesis. Enschede, Netherlands: Cognitive Ergonomics Dept., Twente University, 1998.
8. Waters, K. and Frisbee, J. A coordinated muscle model for speech animation. In *Proceedings of Graphics Interface'95*. Canada, 1995.
9. Weisbrod, R.M. Looking behavior in a discussion group. Unpublished paper, Dept. of Psychology, Cornell University, 1965.