Conversational Gaze Aversion for Virtual Agents

Sean Andrist, Bilge Mutlu, and Michael Gleicher

Department of Computer Sciences, University of Wisconsin–Madison
1210 West Dayton Street, Madison, WI 53706, USA
{sandrist, bilge, gleicher}@cs.wisc.edu

Abstract. In conversation, people avert their gaze from one another to achieve a number of conversational functions, including turn-taking, regulating intimacy, and indicating that cognitive effort is being put into planning an utterance. In this work, we enable virtual agents to effectively use gaze aversions to achieve these same functions in conversations with people. We extend existing social science knowledge of gaze aversion by analyzing video data of human dyadic conversations. This analysis yielded precise timings of speaker and listener gaze aversions, enabling us to design gaze aversion behaviors for virtual agents. We evaluated these behaviors for their ability to achieve positive conversational functions in a laboratory experiment with 24 participants. Results show that virtual agents employing gaze aversion are perceived as thinking, are able to elicit more disclosure from human interlocutors, and are able to regulate conversational turn-taking.

Keywords: Gaze aversion, virtual agents, conversational behavior, intimacy, disclosure, turn-taking

1 Introduction

Engaging in mutual gaze with others has long been recognized as an important component of successful social interactions. People who exhibit high amounts of mutual gaze are perceived as competent, attentive, and powerful [4]. In the same way, virtual agents that use eye contact to exhibit some degree of mutual attentiveness have been shown to achieve a number of positive social and conversational functions, including building rapport with people [20] and increasing positive perceptions of affiliation [2].

Similarly, engaging in gaze *aversion* in conversation also serves a number of communicative functions. Gaze aversions are used to *signal cognitive effort* [4], *modulate intimacy* [1], and *regulate turn-taking* [14]. While social science literature has highlighted the positive functions of gaze aversion, it does not provide the precise temporal measurements required to synthesize a model of gaze aversion for virtual agents that could achieve these functions.

In this work, we enable virtual agents to use gaze aversions to more effectively engage in conversations with people. We first present an analysis of a video corpus of human dyadic conversations from which we obtained temporal parameters of gaze aversion. From these temporal parameters, we designed a gaze controller

Fig. 1. The four agents used in our experiment: Norman, Jasmin, Lily, and Ivy. Norman, Jasmin, and Lily are performing gaze aversions in different directions, while Ivy is maintaining mutual gaze with her interlocutor.

that can generate appropriately timed gaze aversion behaviors for virtual agents. We also present an experimental evaluation of these behaviors to demonstrate their effectiveness in achieving their intended conversational functions. In this experiment, human participants interacted with four different virtual agents in four conversational tasks, each of which was designed to test a different conversational function of gaze aversion (Figure 1).

2 Background

In this section, we present an overview of relevant social and cognitive science research on human gaze aversion. We then review related work on designing effective gaze behaviors for virtual agents.

2.1 Gaze Aversion in Humans

Previous social science research has identified a number of underlying mechanisms to explain human gaze aversion and the social functions it achieves. One such mechanism relevant to our work is the "cognitive interference hypothesis" [5] [8] [9] [12]. This hypothesis posits that gaze aversions facilitate cognitive activity by disengaging the speaker from the environment and limiting visual inputs. Research to support this hypothesis has shown that mutual gaze significantly interferes with the production of spontaneous speech [5]. Research also shows that forcing oneself to look away from a conversational partner while retrieving information from long-term memory or when planning a response to a challenging question significantly improves performance [12] [18].

Previous research has also shown that eye contact is a significant contributor to the intimacy level of an interaction, such that reducing eye contact can decrease the perceived intimacy of a conversation [4]. For example, people generally engage in less eye contact while responding to embarrassing questions than while responding to less objectionable questions [10]. Other work has examined how topic intimacy and eye contact interact over the course of a conversation [1].

Another primary function of gaze aversion is to facilitate turn-taking. Just as making eye contact while listening can serve as a signal that the conversational floor is requested, breaking eye contact while speaking can serve as a signal that the conversational floor is being held and that the speaker has more to say [21]. Kendon [14] found that speakers often look away from their addressees at the beginning of utterances to claim the speaking turn and then look back toward their addressees at the end of their utterance, yielding the turn.

In this work, we group the social-scientific findings discussed above into three broad conversational functions: the *cognitive*, *intimacy-modulating*, and *turn-taking* functions of gaze aversion. These groupings informed our empirical investigation to develop a more computational understanding of how gaze aversions are temporally employed in conversation.

2.2 Gaze Aversion in Virtual Agents

An agent's gaze behavior plays a key role in achieving rich interactions. Well-designed gaze mechanisms—e.g., shifting gaze at turn boundaries during conversation—result in increased task performance and more positive subjective evaluations [13]. Coordinating the head and eyes to maintain a high degree of attention toward human interlocutors has been shown to increase feelings of affiliation with virtual agents [2]. Poor gaze behavior can be worse than the absence of gaze behavior. The positive effects of having an embodied agent—as opposed to only audio or text—can be completely lost if gaze is very poor or random [11].

Previous work has studied different conversational functions of gaze in human-agent interactions, e.g., the use of gaze in facilitating turn management [7] [17]. Wang and Gratch [20] have shown that a virtual agent that continuously gazes toward a human interlocutor is able to increase perceptions of rapport when the gaze is accompanied by nonverbal indicators of positivity and coordination. Continuous gaze without these accompanying behaviors had a negative social impact. Lee et al. [15] developed a statistical model of quick saccadic eye movements for a virtual agent to employ while speaking and listening. This work does not consider the strategic deployment of longer gaze aversions that can be used to achieve specific interactional goals.

While previous research has explored how agents can use gaze to achieve positive social outcomes, a precise account of when agents should avert their gaze from human conversational partners and what social functions these aversions might achieve is still needed. Our work seeks to address this knowledge gap from both theoretical and empirical perspectives through the application of existing social-scientific knowledge and a study of human dyadic conversations to design gaze aversion behaviors for virtual agents.

3 Interaction Design

As outlined above, research in the social sciences has identified a number of conversational functions of gaze aversion. To extend this knowledge to include temporal patterns that will be directly implemented on virtual agent systems, we collected video data from 24 dyadic conversations and derived statistical parameters for the length, timing, and frequency of gaze aversions in relation to speech and conversational functions. We addressed three primary conversational functions of gaze aversion in this analysis, which are defined and described below.

Cognitive – These gaze aversions serve to disengage a speaker's attention from the face of their interlocutor in order to facilitate thinking and remembering [12]. With these aversions, people signal that cognitive processing is occurring while creating an impression that deep thought or creativity is being undertaken [4].

Intimacy-modulating – Gaze aversions also serve to moderate the overall intimacy level of the conversation. Periodic gaze aversions while listening can serve to make speakers more comfortable and reduce negative perceptions associated with staring [1].

Turn-taking – These gaze aversions serve to regulate conversational turn-taking. By looking away at the beginning of an utterance, the speaker strengthens his or her claim over the speaking turn. Looking away during a pause in speech indicates that the conversational turn is being held and that the speaker should not be interrupted [14].

3.1 Data Collection & Analysis

We recruited 24 females and 24 males, aged 18 to 28 and previously unacquainted, for our study. Each dyad engaged in a structured conversation for approximately five minutes. One participant was instructed to learn about the other participant's taste in movies, with the goal of making a movie recommendation. We counterbalanced all conversations for both gender—female and male—and conversational role—recommender and recommendee. We also counterbalanced gender concordance—there was an equal number of gender-matched and gender-mismatched dyads.

Using VCode,[1] we analyzed the recorded videos of the participants' gaze and speech. Video coding was carried out by two independent coders with partial overlap. Sequences of time spent speaking and averting gaze were annotated. Cognitive events were marked as discrete points in time where the participants appeared to be thinking or remembering, commonly occurring at the beginning of responses to questions.

Gaze aversions were coded for the conversational function that they were perceived to be supporting: cognitive, intimacy-modulating, or turn-taking. This coding took place in three passes. In the first pass, the coder was instructed to mark gaze aversions as cognitive if they occurred near labeled cognitive events, e.g., when a participant appeared to be thinking of something new to say. In the second pass, gaze aversions were marked as turn-taking if they occurred near the

[1] http://social.cs.uiuc.edu/projects/vcode.html

Table 1. Gaze aversion parameters in relation to conversational functions and coordinated with (before, after, or within) speech and cognitive events.

Conversational Function	Coordinated With	Parameter	Value
Cognitive	Cognitive Event	Length (sec)	3.54 ($SD = 1.26$)
		Start time (sec)	1.32 before ($SD = 0.47$)
		End time (sec)	2.23 after ($SD = 0.63$)
Intimacy	Speaking	Length (sec)	1.96 ($SD = 0.32$)
		Between (sec)	4.75 ($SD = 1.39$)
	Listening	Length (sec)	1.14 ($SD = 0.27$)
		Between (sec)	7.21 ($SD = 1.88$)
Turn-taking	Utterance Start	Frequency (%)	73.1
		Length (sec)	2.30 ($SD = 1.10$)
		Start time (sec)	1.03 before ($SD = 0.39$)
		End time (sec)	1.27 after ($SD = 0.51$)
	Utterance End	End time (sec)	2.41 before ($SD = 0.56$)

beginning of a speaking turn and were not previously labeled as cognitive. In the third pass, all remaining gaze aversions were labeled as intimacy-modulating. An inter-rater reliability analysis showed substantial agreement on the identification of gaze aversions and their conversational function (Cohen's $\kappa = .747$).

From our analysis, we obtained timing statistics for different kinds of gaze aversions, including the frequency, length, and temporal placement of these gaze aversions relative to speech (Table 1). We also labeled each gaze aversion for its direction, categorized as *up*, *down*, and *side* (Table 2).

3.2 Designing Gaze Aversion for Virtual Agents

Findings from the data analysis were synthesized into a gaze controller for virtual agents that automatically plans and performs gaze aversions to accomplish the conversational functions previously discussed. This controller takes as inputs the current conversational state, the start time and length of upcoming planned utterances, and the time of upcoming cognitive events, and then supplies as outputs the start and end times of planned gaze aversions to be executed by the agent. The exact timings of the gaze aversions are drawn from the parameter distributions shown in Table 1. These distributions are modeled as Gaussian functions in the current implementation.

Source of inputs – Recognized speech from the user is passed to a dialogue manager that associates a semantic tag with the utterance and plans the agent's speech accordingly. For example, if the dialogue manager receives a recognized question, it will produce the associated answer. The dialogue manager sends

Table 2. Frequency of gaze aversions up, down, and to the side for each conversational function.

Conversational Function	**Frequency Up**	**Frequency Down**	**Frequency Side**
Cognitive	39.3%	29.4%	31.3%
Intimacy-modulating	13.7%	28.8%	57.5%
Turn-taking	21.3%	29.5%	49.2%

upcoming cognitive events, speech events, and the current conversational state to the gaze controller. Cognitive events could alternatively be passed to the gaze controller from a dedicated cognitive architecture, but in our implementation, cognitive events were created by labeling some of the agent's utterances as "cognitively difficult" and generating a cognitive event at the beginning of those utterances.

Gaze controller – Cognitive events are represented with a single timestamp, t_c. Planned speech events are represented as a vector containing start and end times, $[t_s, t_e]$. Conversational state, CS, indicates that the agent is currently in either *speaking* or *listening* mode. As the gaze controller receives these inputs from the dialogue manager, it continuously plans future gaze aversions in real-time. The first priority is to plan gaze aversions around upcoming cognitive events, t_c. The start and end times of the gaze aversion, $[GA_s, GA_e]$, are computed by drawing from the cognitive parameter distributions shown in Table 1. The controller next looks for upcoming speech events and calculates first if a turn-taking gaze aversion will be performed. If a gaze aversion will be performed, the controller then calculates $[GA_s, GA_e]$ around the start of the utterance, t_s, by drawing from the turn-taking parameter distributions provided in Table 1. Finally, the controller calculates the next intimacy gaze aversion according to CS. These gaze aversions are only planned for times when cognitive and turn-taking aversions are not already planned. Also, intimacy gaze aversions are prohibited near the end of utterances, t_e, so that virtual agents can appropriately pass the floor by maintaining mutual gaze.

Example simulation – Figure 2 illustrates a simulation of the gaze aversion behaviors produced by our controller. In this example, two agents, A1 and A2, are having a conversation. Both are using the gaze aversion controller. A1 asks a question constructed from two utterance parts with a pause in between. A turn-taking gaze aversion is planned and executed around the start of the second utterance in order to hold the conversational floor. While A1 is listening, it occasionally looks away to regulate the intimacy of the conversation. Upon recognizing A1's question, A2 plans to give its response, which has been tagged with a cognitive "thinking" event at its beginning. The gaze controller plans and executes a cognitive gaze aversion around the beginning of the utterance to express this thinking. All other gaze aversions in the example have been similarly produced by the controller to achieve one of the three conversational functions.

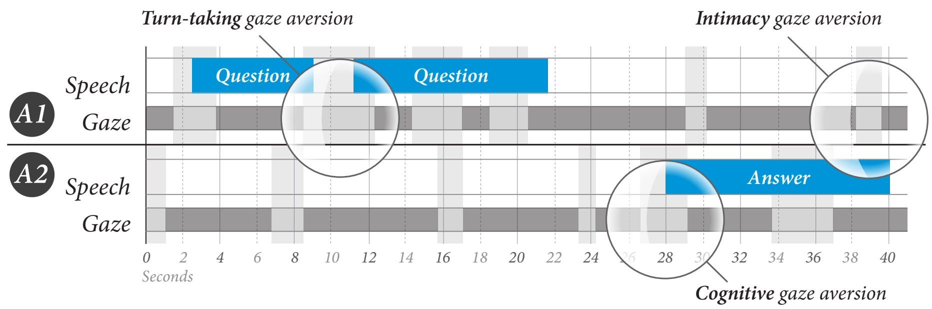

Fig. 2. Gaze aversions created by our controller for two agents in conversation. Dark gray intervals on the gaze stream indicate periods of gazing toward the interlocutor, and light gray intervals indicate gaze aversions.

4 Experimental Evaluation

We developed four hypotheses to test how agents might use the gaze aversion behaviors generated by our controller to achieve conversational functions. The first two hypotheses focus on the cognitive function, the third on the intimacy-modulation function, and the fourth on the turn-taking function.

Hypothesis 1 – A virtual agent averting its gaze while not currently speaking will be perceived as *thinking*, whereas an agent that does not avert its gaze will not elicit this impression.

Hypothesis 2 – Virtual agents that display gaze aversions at the start of utterances will be rated as being more *thoughtful* and creative than virtual agents that do not display gaze aversions.

Hypothesis 3 – Virtual agents that display periodic gaze aversions while listening will increase a human interlocutor's comfort and elicit more *disclosure* than agents that do not display gaze aversions.

Hypothesis 4 – Virtual agents that display gaze aversions during pauses will be perceived as *holding the floor* and will be interrupted less than agents that do not display gaze aversions.

4.1 Study Design

Twenty-four participants were recruited for this study (12 females and 12 males), aged between 18 and 45 ($M = 23$, $SD = 6.82$). All participants were native English speakers and were recruited from the University of Wisconsin–Madison campus.

The experiment involved a single independent variable, *gaze aversion condition*, with three conditions varying between participants. One condition involved the virtual agents using gaze aversions generated by the controller described in the previous section, which we call the *good timing* condition. The other two conditions were baselines for comparison. The first baseline was a *static gaze* condition in which the virtual agents did not employ any gaze aversions. The second baseline was a *bad timing* condition in which the virtual agent employed just as many gaze aversions as in the *good timing* condition but with reverse timings. When the gaze controller indicated that a gaze aversion should be made, the *bad timing* model engaged a mutual gaze shift, and vice versa. This third condition was included as a baseline to show that both the presence and the timing of gaze aversions are important for achieving positive social outcomes.

We created separate tasks to test each hypothesis, each using a different virtual agent (Figure 1). Participants were randomly assigned to one of the three gaze aversion conditions, which was held constant for all four tasks (8 participants per condition). Tasks were presented in random order.

Task 1 – The first task was designed to test Hypothesis 1. The participant was told that the virtual agent, Norman, was training to work at a help desk in a campus library. The participants were given five library-related questions to ask Norman. They were instructed to ask each question and listen to the response. Norman would pause for 4 to 10 seconds (randomly determined) before answering each question. Participants were instructed to ask a question again if they thought Norman did not understand or was not going to answer. The primary measure was the time participants waited for Norman to respond to questions before interrupting him to ask the question again. For this task, we deliberately chose an agent with an abstract design that minimally elicits attributions of intent or thought in order to ensure that the agent's gaze aversions were solely responsible for the impression of thinking, unconfounded from any other animation variables.

Task 2 – The second task was designed to test Hypothesis 2. For this task, participants were instructed to ask the agent, Jasmin, a series of five common job interview questions. Jasmin was programmed to respond with answers taken from real-world job interviews. Participants rated each response immediately after it was given on four seven-point rating scales. These scales measured the perceived thoughtfulness, creativity, disclosure, and naturalness of each response. In our analysis, we combined the scales into a single broad indicator of *thoughtfulness*. Internal consistency was excellent for this measure (Cronbach's $\alpha = .903$).

Task 3 – The third task was designed to test Hypothesis 3. In this task, participants spoke to an agent named Lily, who was introduced as training to be a therapist's aide who would conduct preliminary interviews with incoming clients. Lily asked the participant a series of five questions of increasing intimacy, and participants were instructed to respond with as much or as little detail as they wished. Questions ranged in intimacy, from "What do you like to do in your free time?" to "What is something you would like to accomplish before dying?" The primary measure for the third task was the *degree of self-disclosure*, specifically the breadth of disclosure. Breadth of disclosure was obtained using a word count of participants' responses to Lily's questions. Word count has been validated as an appropriate measure of disclosure in previous research on how computers can be used to elicit self-disclosure from people [16].

Task 4 – The fourth task was designed to test Hypothesis 4. Participants were provided with a list of five questions to ask a virtual agent named Ivy, with the goal of getting to know each other. Participants were instructed to ask each

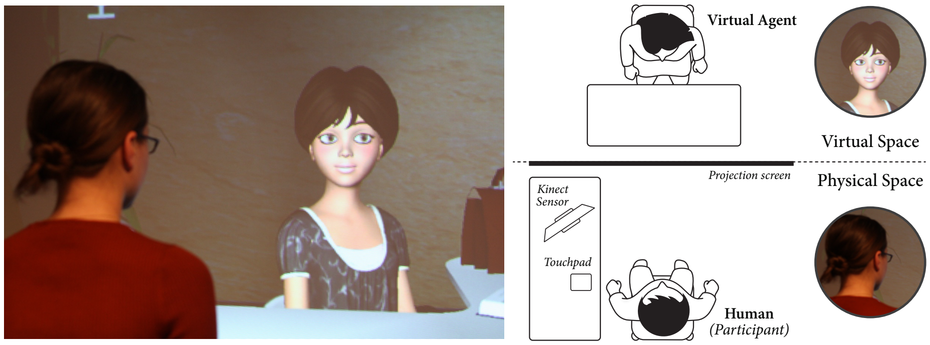

Fig. 3. An experimenter demonstrating the interaction with the virtual agent on a life-size projected display (left) and the physical setup of the experiment (right).

question, listen to Ivy's response, and then reciprocate with their own response to the same question. Ivy's responses had two parts, separated by a pause between 2 and 4 seconds in length (randomly determined). If participants started speaking during the pause, Ivy refrained from giving the second part of her response. The primary measure of the fourth task was the time participants waited for Ivy to be silent during the pause in her speech before interrupting.

4.2 Setup & Procedure

The experiment was implemented using a custom character animation framework built on top of the Unity game engine.[2] The agent's behaviors were implemented as Unity scripts. In the second, third, and fourth tasks, the agents periodically smiled, blinked, and nodded their heads to achieve greater naturalness and humanness. In all tasks and conditions, the agent's eyes made small, periodic saccadic motions according to the model presented by Lee et al. [15]. The agents were created using commercially available parametric base figures. Audio and lip-sync animations were pre-recorded.

Gaze aversions were executed using the head-eye coordination model described by Andrist et al. [3] with a moderate amount of head movement. Head alignment was high as the agent oriented its gaze back to the interlocutor, in accordance with the finding that high head alignment increases people's feelings of affiliation with agents [2].

After giving informed consent, the experimenter led each participant into the study room and gave a brief introduction to the experiment. The participant sat in a chair approximately six feet away from a large screen on which the life-size virtual agent was projected (Figure 3). A wireless touchpad was used as a button to begin each conversational task, and a Kinect microphone was used for capturing speech. The Microsoft Speech Platform[3] was used for speech recognition in combination with a custom dialogue manager specific to each

[2] http://www.unity3d.com
[3] http://msdn.microsoft.com

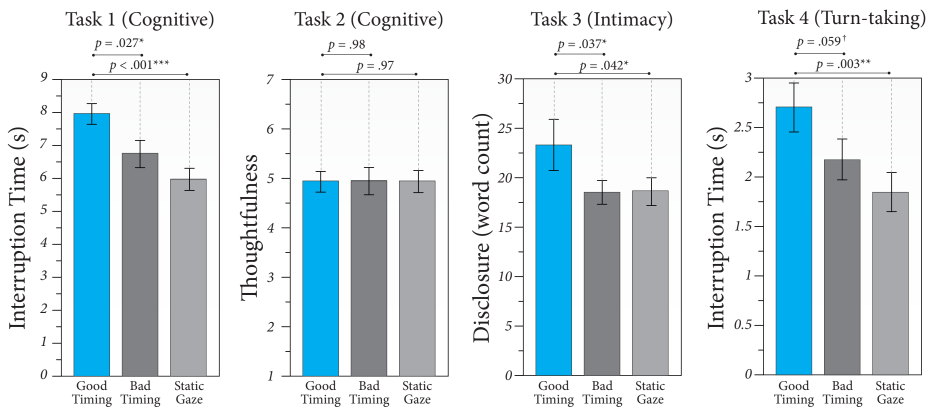

Fig. 4. The results of the evaluation. Virtual agents that displayed gaze aversions with appropriate timings successfully conveyed the impression that they were "thinking," elicited more disclosure from participants, and were better able to hold the conversational floor during breaks in speech. (†), (*), (**), and (***) denote $p < .10$, $p < .050$, $p < .010$, and $p < .001$, respectively.

task. After completing all four tasks, the participant responded to a survey of demographic characteristics and was debriefed. The study took approximately 30 minutes, and each participant was given $5 as compensation.

4.3 Results

We performed a mixed-design analysis of covariance (ANCOVA) to assess how agent gaze aversion behaviors affected the dependent variable for each task. Participant gender was included as a covariate to control for gender differences. Question ID was included as a covariate to control for learning effects. Planned comparisons were carried out as *apriori* contrast tests using Scheffé's method.

Hypothesis 1 – Our analysis supported this hypothesis. The time given to the virtual agent before interrupting was significantly higher when the agent used proper gaze aversion with good timing rather than bad timing, $F(1, 110) = 5.06, p = .027$, or with no gaze aversion at all, $F(1, 110) = 12.71, p < .001$.

Hypothesis 2 – Our analysis did not support this hypothesis. Participants' ratings did not differ for virtual agents using proper gaze aversion over agents using gaze aversion with bad timing, $F(1, 110) = 0.0004, p = .98$, or with no gaze aversion at all, $F(1, 110) = 0.002, p = .97$.

Hypothesis 3 – Our analysis supported this hypothesis. Virtual agents using gaze aversions with good timing elicited significantly more disclosure from participants than when their gaze aversions were badly timed, $F(1, 110) = 4.48, p = .037$, or when they used no gaze aversion, $F(1, 110) = 4.25, p = .042$.

Hypothesis 4 – Our analysis partially supported this hypothesis. The time given to the virtual agent during its pause before interrupting was marginally higher when the agent used properly-timed gaze aversion than when its gaze aversions were badly timed, $F(1, 110) = 3.64, p = .059$, and significantly higher than when it did not use gaze aversion at all, $F(1, 110) = 9.48, p = .003$. All of our primary results are illustrated in Figure 4.

5 Discussion

Virtual agents that displayed gaze aversion behaviors generated by our controller were partially successful in achieving the cognitive conversational function of gaze aversions. As shown in Task 1, virtual agents successfully used gaze aversion to indicate that they were engaged in a form of cognitive processing with a response forthcoming and thus delayed interruptions by a human interlocutor. However, as shown in Task 2, using gaze aversions before responses did not affect how thoughtful participants thought those responses were. A possible explanation for this result is that while participants respond *behaviorally* to an agent using gaze aversion to achieve conversational functions, these cues fail to elicit explicit attributions of thought when participants are asked to reflect on the interaction afterwards.

Virtual agents displaying gaze aversion behaviors generated by our controller were successful in eliciting more disclosure from participants. Measurements of the breadth of participants' responses in Task 3 show that participants disclosed more when the virtual agent periodically looked away from them with appropriate timings than when the agent did not look away or looked away at inappropriate times.

Finally, virtual agents displaying gaze aversion behaviors generated by our controller were successfully able to regulate conversational turn-taking. By averting their gaze at the appropriate time in Task 4, virtual agents more effectively held the conversational floor than when they used gaze aversion at inappropriate times or not at all.

Designers must consider gaze aversion as more than "lack of eye contact" and instead as a powerful cue that can achieve conversational goals. If the goal is to elicit disclosure from a human, the virtual agent should use gaze aversion to regulate the intimacy of the conversation. When virtual agents need to pause in their speech, e.g., to process information or plan their next utterance, gaze aversion is an effective strategy to hold the conversational floor and indicate to the human that a new utterance is forthcoming. This idea is similar to work by Shiwa et al. [19], which showed that robots can use conversational fillers to successfully alleviate users' negative perceptions to long system response times.

5.1 Limitations & Future Work

Although the gaze aversion strategies presented in this paper are closely tied to the conversational states of speaking and listening, future work should concentrate on connecting gaze aversions more closely with the content and structure of speech. Previous research by Cassell et al. [7] identified relationships between gaze behavior and information structure of utterances, specifically the theme and rheme of sentences. Integrating these findings with our gaze aversion controller would be a useful extension to the current work.

A limiting assumption of our controller is that the gaze aversion behaviors generated are stable over time, while in reality these behaviors likely change over the course of a conversation due to increasing familiarity with the interlocutor, changing emotions and level of comfort, and so on. In future work, we plan to develop models of gaze that dynamically adjust gaze aversion strategies over time as well as retain the significant edge cases of behavior that are potentially lost by our current statistical approach of collapsing data into averaged distributions.

Another limitation of our work is that the gaze aversion behaviors of the virtual agent do not take into account the gaze behavior of the user. By tracking the gaze of the user, a virtual agent could more effectively modulate the amount of mutual gaze exhibited in the interaction in order to better regulate intimacy. It could also assess whether it has the attention of the user before attempting nonverbal behaviors that have an associated conversational goal. Previous research has explored the development of interactive gaze models for virtual agents, such as work by Bee et al. [6]. Future work might develop interactive models of gaze *aversion* that more dynamically employ aversion behaviors in human-agent conversations.

6 Conclusion

Gaze aversions are commonly associated with negative social outcomes, including discomfort, inattention, and deceit, but in reality they serve a number of important positive conversational functions, including cognitive, intimacy-modulating, and turn-taking functions. In this paper, we demonstrated how to enable virtual agents to use gaze aversions to achieve these functions in conversations with people. We presented an analysis of human dyadic conversations that informed the development of a gaze aversion controller that can automatically plan and execute appropriately timed gaze aversions for virtual agents. We also presented an experiment that evaluated the gaze aversion behaviors generated by the controller for their effectiveness in achieving positive conversational functions. The experiment demonstrated that virtual agents using gaze aversions generated by our controller were perceived as thinking, elicited more disclosure from human interlocutors, and effectively managed turn-taking. Our findings suggest that gaze aversion is a powerful conversational cue that designers should draw on in order to create effective and natural human-agent interactions.

Acknowledgments

This research was supported by National Science Foundation award 1017952. We would like to thank Faisal Khan, Brandi Hefty, Allison Terrell, Danielle Albers, Irene Rae, and Brandon Smith for their help in data collection, video coding, and providing voices for our virtual agents.

References

1. A. Abele. Functions of gaze in social interaction: Communication and monitoring. *Journal of Nonverbal Behavior*, 10(2):83–101, 1986.
2. S. Andrist, T. Pejsa, B. Mutlu, and M. Gleicher. Designing effective gaze mechanisms for virtual agents. In *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems*, pages 705–714. ACM, 2012.
3. S. Andrist, T. Pejsa, B. Mutlu, and M. Gleicher. A head-eye coordination model for animating gaze shifts of virtual characters. In *Proceedings of the 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction*, pages 4:1–4:6. ACM, 2012.
4. M. Argyle and M. Cook. *Gaze and mutual gaze.* Cambridge University Press Cambridge, 1976.
5. G. W. Beattie. A further investigation of the cognitive interference hypothesis of gaze patterns during conversation. *British Journal of Social Psychology*, 20(4):243–248, 1981.
6. N. Bee, J. Wagner, E. André, T. Vogt, F. Charles, D. Pizzi, and M. Cavazza. Discovering eye gaze behavior during human-agent conversation in an interactive storytelling application. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*, pages 9:1–9:8. ACM, 2010.
7. J. Cassell, O. Torres, and S. Prevost. Turn taking vs. discourse structure: How best to model multimodal conversation. *Machine conversations*, pages 143–154, 1999.
8. G. Doherty-Sneddon and F. Phelps. Gaze aversion: A response to cognitive or social difficulty? *Memory & Cognition*, 33(4):727–733, 2005.
9. H. Ehrlichman and D. Micic. Why do people move their eyes when they think? *Current Directions in Psychological Science*, 21(2):96–100, 2012.
10. R. Exline, D. Gray, and D. Schuette. Visual behavior in a dyad as affected by interview content and sex of respondent. *Journal of Personality and Social Psychology*, 95:201–209, 1965.
11. M. Garau, M. Slater, S. Bee, and M. Sasse. The impact of eye gaze on communication using humanoid avatars. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 309–316. ACM, 2001.
12. A. M. Glenberg, J. L. Schroeder, and D. A. Robertson. Averting the gaze disengages the environment and facilitates remembering. *Memory & Cognition*, 26(4):651–658, 1998.
13. D. Heylen, I. Van Es, E. Van Dijk, A. Nijholt, J. van Kuppevelt, L. Dybkjaer, and N. Bernsen. Experimenting with the gaze of a conversational agent. In *Proceedings of the International CLASS Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems*, pages 93–100. Kluwer Academic Publishers, 2005.
14. A. Kendon. Some functions of gaze-direction in social interaction. *Acta psychologica*, 26(1):22–63, 1967.
15. S. Lee, J. Badler, and N. Badler. Eyes alive. In *ACM Transactions on Graphics (TOG)*, pages 637–644. ACM, 2002.
16. Y. Moon. Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of Consumer Research*, 26(4):323–339, 2000.
17. C. Pelachaud and M. Bilvi. Modelling gaze behavior for conversational agents. In *Intelligent Virtual Agents*, pages 93–100. Springer, 2003.
18. F. G. Phelps, G. Doherty-Sneddon, and H. Warnock. Helping children think: Gaze aversion and teaching. *British journal of developmental psychology*, 24(3):577–588, 2006.

19. T. Shiwa, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita. How quickly should communication robots respond? In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 153–160. IEEE, 2008.
20. N. Wang and J. Gratch. Don't just stare at me! In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2010.
21. J. M. Wiemann and M. L. Knapp. Turn-taking in conversations. *Journal of Communication*, 25(2):75–92, 1975.