Conversational Gaze Aversion for Humanlike Robots

Sean Andrist, Xiang Zhi Tan, Michael Gleicher, and Bilge Mutlu
Department of Computer Sciences, University of Wisconsin–Madison
1210 West Dayton Street, Madison, WI 53706 USA
{sandrist, xtan, gleicher, bilge}@cs.wisc.edu

ABSTRACT

Gaze aversion—the intentional redirection away from the face of an interlocutor—is an important nonverbal cue that serves a number of conversational functions, including signaling cognitive effort, regulating a conversation's intimacy level, and managing the conversational floor. In prior work, we developed a model of how gaze aversions are employed in conversation to perform these functions. In this paper, we extend the model to apply to conversational robots, enabling them to achieve some of these functions in conversations with people. We present a system that addresses the challenges of adapting human gaze aversion movements to a robot with very different affordances, such as a lack of articulated eyes. This system, implemented on the NAO platform, autonomously generates and combines three distinct types of robot head movements with different purposes: face-tracking movements to engage in mutual gaze, idle head motion to increase lifelikeness, and purposeful gaze aversions to achieve conversational functions. The results of a human-robot interaction study with 30 participants show that gaze aversions implemented with our approach are perceived as intentional, and robots can use gaze aversions to appear more thoughtful and effectively manage the conversational floor.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems—*human factors, software psychology*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*evaluation/ methodology, user-centered design*

General Terms

Design, Experimentation, Human Factors

1. INTRODUCTION

Gaze is an important nonverbal behavior for social interaction [9]. Research in human-robot interaction (HRI) has investigated the role of gaze behavior across a variety of settings, including open-ended conversation [11], presenting information [16], and storytelling [19]. A recent survey of research on social gaze in HRI proposes a computationally oriented definition of social gaze as a mapping from social functions to the expression of one or more discrete gaze actions [27]. In this paper, we focus on one of these actions—*gaze aversion*—and show they might enable humanlike robots to achieve a number of positive social functions in conversations with humans.

HRI'14, March 3–6, 2014, Bielefeld, Germany.

ACM 978-1-4503-2658-2/14/03 ...$15.00.
http://dx.doi.org/10.1145/2559636.2559666.

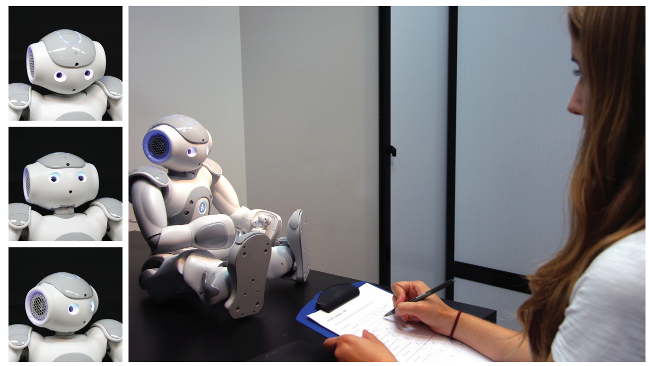

Figure 1: A human conversational partner interacting with the NAO robot. Three example gaze aversion directions implemented for the NAO are shown: down, up, and to the side.

Gaze aversion is defined as the intentional redirection of gaze away from the face of an interlocutor, and it is used in conversations to achieve three primary functions: cognitive, intimacy modulation, and floor management. First is the *cognitive* function: speakers spend much more time averting their gaze than listeners in order to better attend to the planning and delivery of their utterances while limiting external distraction [4]. Second is the *intimacy-modulating* function: periodic gaze aversions while speaking or listening can serve to modulate the overall level of intimacy in the conversation [1]. Third is the *floor management* function: looking away while pausing during speech is used to indicate that the conversational floor is being held and the speaker should not be interrupted [13]. These three functions—cognitive, intimacy, and floor management—correspond with three social contexts identified in previous work on social gaze in HRI: projecting mental state, establishing agency, and regulating the interaction process respectively [27].

While the social science literature has highlighted the positive functions of gaze aversion, it does not provide the exact measurements required to synthesize a model of gaze aversion for robots that could achieve these functions. In previous work [3], we analyzed a corpus of human-human conversations to obtain precise spatial and temporal parameters of gaze aversion movements in relation to speech and conversational function. This analysis informed the design of a gaze controller for virtual agents that generated appropriately timed gaze aversions during conversations with people. However, the prior work does not consider the challenges of applying the model to an autonomous robot, nor does it provide any evidence that gaze aversions can be effectively used by robots.

In this paper, we demonstrate that humanlike robots can use gaze aversions to achieve conversational goals. We apply techniques that address the challenges of applying a human-based gaze aversion model to a robot, describe an implementation on the NAO platform, and evaluate the resulting system in a human-robot interaction study. Our work considers three key technical challenges: adapting the human movements to a robot with non-human affordances, making the movements appear lifelike and intentional, and integrating the gaze aversion movements with other head movements. A description of the system, along with a link to downloadable source code, is presented in Section 3.2. In our experimental evaluation, presented in Section 4, human participants interacted with two robots in four conversational tasks (Figure 1). Through this evaluation, we show how the gaze aversions employed by the robot were perceived as intentional, served to make the robot appear more thoughtful, and helped the robot manage the conversational floor.

2. BACKGROUND

In this section, we present an overview of social and cognitive science research on human gaze aversion in conversations, in which three primary functions of gaze aversion have been identified: facilitation of cognitive processing, intimacy modulation, and floor management. We then review related work on designing effective gaze mechanisms for robots.

2.1. Conversational Gaze Aversion

The frequency of speaker gaze aversions in conversation has been shown to be related to the difficulty of *cognitive processing* [10]. When constant mutual gaze is required from someone speaking spontaneously in social interaction, that person's speech becomes significantly impaired [7]. Averting gaze has the practical benefit of improving cognitive performance. This is because an interlocutor's face is rich in social information, and is a cognitively demanding visual target. When people avert their gaze from their interlocutor, they are able to deploy additional cognitive resources to the task of thinking or remembering information. With these aversions, a speaker signals to the listener that cognitive processing is occurring, creating the impression that deep thought or creativity is being undertaken in formulating their speech [4].

Eye contact is a primary contributor to the level of *intimacy* in a conversation along with physical proximity, topic intimacy, amount of smiling, and so on [5]. If one of these dimensions of intimacy is disturbed, compensatory changes will likely occur along the other dimensions. For example, research has shown that children avert their gaze more when answering questions in face-to-face rather than video-mediated conversations [8]. In general, people frequently avert their gaze to alleviate feelings of self-consciousness and, while listening, to make speakers more comfortable and to reduce negative perceptions associated with staring [1].

Gaze is important for facilitating turn-taking and conversational *floor management* [13]. When exchanging speaking turns in conversation, a pattern is frequently observed in which the first speaker finishes speaking, looks toward their interlocutor and engages in momentary mutual gaze, and finally the second speaker averts their gaze and begins their speaking turn [21]. 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 is also used to indicate that the conversational floor is being held and that the speaker should not be interrupted [13].

2.2. Robot Gaze Mechanisms

Previous research has shown that robots can greatly benefit from the utilization of humanlike gaze mechanisms [16, 19, 20, 28]. Modeling humanlike gaze mechanisms enables conversational robots to signal different participant roles to human interlocutors, manage turn-exchanges, and shape how users perceive the robot and the conversation [20]. Human comprehension of robot speech is improved when a robot gazes to objects it is speaking about in a similar way and with similar timings as found in human gaze behavior [28]. Robots that use gaze cues are perceived as possessing mental states and intentionality, as evidenced by previous work on gaze leakage in a human-robot guessing game [20].

To be truly effective, a robot's gaze mechanisms need to be employed contingently based on the behaviors of its human interlocutors. A robot that gazes responsively toward a human user is capable of eliciting a stronger feeling of being looked at than a robot that uses non-responsive—i.e., static or random—gaze [30]. In recent work, HRI researchers have developed a responsive, rule-based system for generating robot head nods, tilts, and gazes based on discrete dialogue acts in conversation, such as those associated with turn-taking, backchannels, and conversational fillers [16]. When employing gaze aversions in conversation, robots should do so responsively to the speech behavior of its human interlocutor.

In previous work [3], we showed how to enable virtual agents to effectively use gaze aversions responsively in conversations with people. However, it is not obvious that these results extend to HRI, as the prior work did not consider the technical challenges specific to robots. Robots and virtual agents differ along a number of social dimensions, including realism, social presence, lifelikeness, and physical proximity [24]. Several studies have demonstrated effects of these differences [6, 14, 24].

A fundamental difference between the virtual agents we used in previous work [3] and the robots used in the current work is the difference in physical affordances available for carrying out gaze motions. Overall differences in geometry mean that gaze motions and control laws need to be adapted [22]. Even more critically, the robots lack articulated eyes and must rely solely on head motions to convey gaze motions, making it unclear if they are capable of eliciting the same positive conversational outcomes found in the virtual agent work [3]. Previous research points to the possibility of such capabilities, including work which has shown that people are capable of recognizing a robot's gaze according to its head orientation [11] and that robots can use head motions alone to gaze effectively in a storytelling scenario [19].

3. ROBOT GAZE AVERSION

In prior work [3], we derived a gaze aversion model consisting of precise spatial and temporal parameters—including length, timing, and frequency—in relation to conversational functions and speech. In the next section, we review the data collection and model parameters relevant for the design of a robot gaze controller. Following that, we present our design of gaze aversion motions for humanlike robots without articulated eyes, secondary head motions for achieving mutual gaze and lifelikeness, techniques for combining the different head motions, and the overall system implementation for expressing conversational gaze aversion on the NAO platform.

3.1. Modeling Gaze Aversion

We recruited 24 females and 24 males, aged 18 to 28 and previously unacquainted, for data collection (Figure 2). Each dyad engaged in a structured conversation for approximately five minutes. We analyzed recorded videos of the interactions for 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.

Gaze aversions were coded for the conversational function that they were perceived to be supporting: cognitive, intimacy-modulating, or floor management. This coding took place in three passes. In the first pass, the coder marked gaze aversions as cognitive if they occurred near perceived cognitive events, e.g., when a participant appeared to be thinking of a response to a question. In the second pass, gaze aversions were marked as floor management

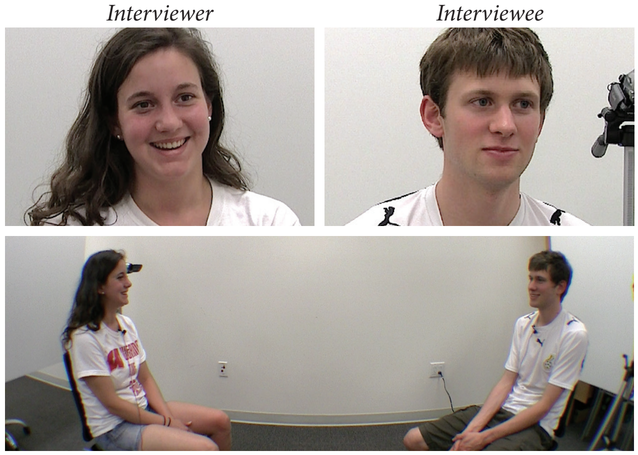

Figure 2: A participant dyad from the data collection. The participants were designated as the *interviewer* and the *interviewee*. The interviewers were instructed to ask interviewees about their movie preferences.

if they occurred near the beginning of a speaking turn or during a pause in speech. 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. Each of these parameters is modeled as a Gaussian distribution with means and standard deviations derived from the data. For cognitive gaze shifts, we model the length ($M = 3.54s$, $SD = 1.26s$), start time in relation to cognitive events ($M = -1.32s$, $SD = 0.47s$), and end time after cognitive events ($M = 2.23s$, $SD = 0.63s$). A cognitive event is any point in time at which the robot should display a state of deep cognitive processing, e.g., at the beginning of a response to a user's question. For intimacy-modulating gaze aversions, we model the length and time between consecutive gaze aversions while speaking (length: $M = 1.96s$, $SD = 0.32s$; time between: $M = 4.75s$, $SD = 1.39s$), and while listening (length: $M = 1.14s$, $SD = 0.27s$; time between: $M = 7.21s$, $SD = 1.88s$). For floor management gaze aversions, we model the length ($M = 2.30s$, $SD = 1.10s$), start time in relation to the start of the next utterance ($M = -1.03s$, $SD = 0.39s$), and the end time in relation to that same utterance ($M = 1.27s$, $SD = 0.51s$). Our model also captures the time before the end of a floor-passing utterance at which point mutual gaze is engaged with the interlocutor and no more gaze aversions are generated during that utterance ($M = -2.41s$, $SD = 0.56s$). Figure 3 illustrates each type of gaze aversion.

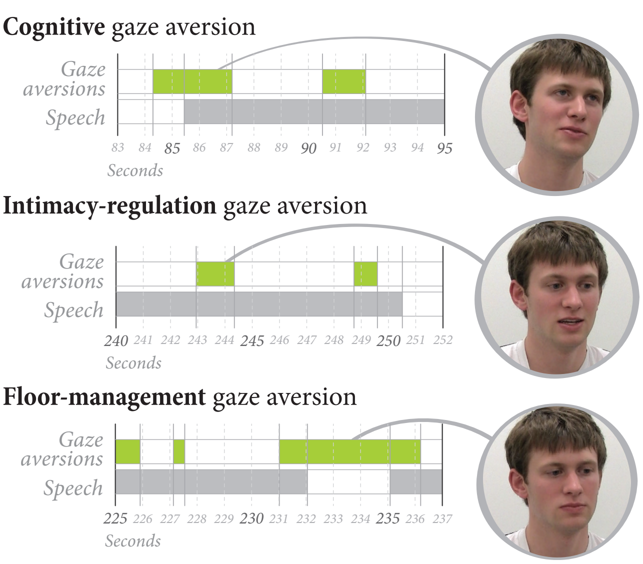

Figure 3: Three examples of gaze aversions from our human-human conversational data. Top: an upward *cognitive* gaze aversion at the beginning of a question response. Middle: a short *intimacy-regulation* gaze aversion while speaking. Bottom: a *floor-management* gaze aversion during a pause.

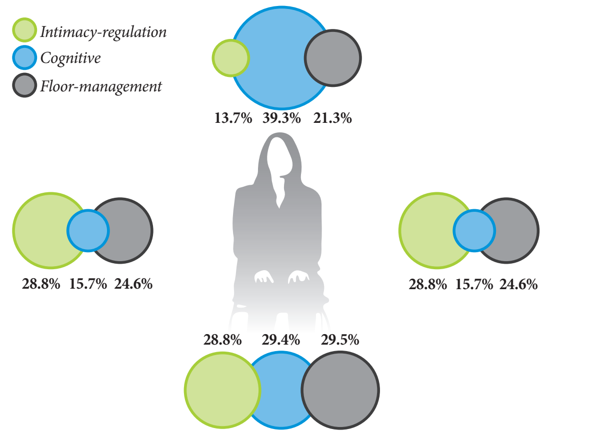

Figure 4: Percentages of gaze aversions directed up, down, and to the side, split by conversational function. Intimacy-regulating and floor-managing gaze aversions are more likely to be directed sideways, while cognitive gaze aversions are more likely to be directed upwards.

We also labeled each gaze aversion for its direction as *up*, *down*, and *side* (Figure 4) and found that intimacy-regulating and floor-managing gaze aversions and cognitive gaze aversions were more likely to be directed sideways and upwards, respectively.

3.2. Robot Gaze Aversion Implementation

In this section, we describe methods to overcome three key technical challenges in adapting our gaze aversion model to a head controller for conversational robots: adapting the human movements to a robot with non-human affordances, making the movements appear lifelike and intentional, and integrating the gaze aversion movements with other head movements. To adapt gaze movements to a robot platform without articulated eyes, we substitute head movement for combined eye and head rotations, extending prior work that showed robot head motions to serve as valid gaze signals [19]. To make the movements appear lifelike and intentional, we combine aversion control with face tracking and structured random movements to create idle motion. To realize these behaviors in an autonomous system, we implement them in a predictive filtering framework that affords graceful combination of multiple goals and effective reaction to external events.

3.2.1. Aversion Movement Design

A robot without articulated eyes must rely on head motion alone to carry out gaze aversion motions. There are two considerations when designing these motions: the magnitude and the dynamics.

The appropriate magnitudes for gaze-averting head motions were determined in an iterative process to achieve a natural subjective appearance. Our goal was to generate head movements that are not too extreme yet clearly serve to avert the robot's gaze away from the user. Our system generates vertical gaze-averting head movements with a magnitude of approximately 20 degrees, with horizontal and downward head movements at 28 and 22 degrees respectively. For intimacy-modulating gaze aversions, these angles are scaled by a factor of 0.4, because these gaze aversions occur quite frequently and were observed in our human-human data to be employed with more subtle eyes-only motions.

For the generated gaze aversions, we attempted to recreate the head velocity profile identified in neurophysiological research on human gaze shifts [15]. This profile resembles standard ease-in/ease-out curves found in animation, in which the head rotation starts slow, speeds up during the bulk of the shift, then slows down before coming to a halt. The velocity profile was implemented using programmatically-defined bezier curves. Care was taken to respect both the upper and effective lower limits of the motors' rotational speeds, ensuring smooth motions throughout the robot's gaze shifts.

3.2.2. *Achieving Mutual & Lifelike Gaze*

In order to interact effectively with humans and accomplish a general feeling of lifelikeness, the robot must use its head for not only gaze aversion, but also mutual gaze. To engage in mutual gaze, the robot must track the user's face. For our system, we use a face-tracking algorithm that used a Microsoft Kinect situated behind the robot. The robot continuously adjusts its head rotation based on the results of the face tracking algorithm, polled every 200*ms*.

At this point the robot is capable of executing gaze aversions and face tracking, however a problem remains in that the transition between statically gazing at the user and engaging in a gaze aversion is still quite abrupt. When the user is sitting still and the robot is not currently engaged in a gaze aversion motion, the robot's head becomes motionless and loses its sense of liveliness. To address this problem, we generate a small amount of idle head motion for the robot to execute at all times. This idle motion was implemented by adding a small amount of structured noise, generated by a Perlin noise function [23], to the results of the face tracking algorithm. A Perlin noise function generates band-limited, pseudorandom signals that are useful for emulating biological forms and motions.

3.2.3. *Combining Behaviors*

The robot uses its head for three different types of gaze motions: gaze aversions, face tracking, and idle motion. At any point in time, the robot has multiple distinct goals for its target head rotation. These goals must be combined in a natural manner that maintains an ability to be reactive to new goals. To solve this problem, we use a Kalman filter [12], a linear predictive filter used for estimating the state of a system given past states and target goals. The filter predicts the appropriate motion by blending estimated trajectories generated from the current state and current goals, and corrects these estimates as the goals change. The filter gains were chosen empirically to provide an appropriate balance between smoothness and reactivity.

Target head rotations from the three types of motions are streamed through this filter and combined into a single head rotation signal, allowing for graceful transitions between motions. Without this filter, head motions would have to be serially generated and completed, e.g., a head tracking motion to the face of the user would have to be completed before a gaze aversion motion could be executed. The Kalman filter solves this problem by blending estimated trajectories over time, resulting in smooth, interruptible motions.

3.2.4. *System Integration*

The target platform for this work was the NAO, a programmable humanoid robot manufactured by Aldebaran Robotics. [1] The NAO has 25 degrees of freedom, including two in its neck, and a multitude of sensors. All behaviors for the NAO were implemented using the .NET SDK provided by Aldebaran. Gaze aversion behaviors were implemented on the NAO using two controllers: a high-level *gaze controller* to plan the timing and direction of gaze aversions and a *head controller* to physically execute the gaze shifts. The source code for the entire system has been made available online. [2]

The head controller is situated at the end of an overall system pipeline that also includes a speech recognition system and a dialogue manager. A Microsoft Kinect with the Microsoft Speech Platform [3] are utilized for speech recognition. 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. Robot speech is generated in our system by playing pre-recorded audio files, but the NAO's built-in text-to-speech system can also be used.

The dialogue manager sends upcoming speech events and the current conversational state to the gaze controller. As the gaze controller receives these inputs, it continuously plans gaze aversion motions to be executed by the head controller. The exact timings of the gaze aversions are drawn from the parameter distributions reported in the previous section and visually depicted in Figure 3. The direction of each gaze aversion is similarly generated according to its function and the directional likelihoods presented in Figure 4.

Also included in the overall system are a wireless touchpad and the NAO's chest light. The wireless touchpad is used by the user to signal that the NAO can begin the next phase of the interaction, depending on the context. For example, after the user is done giving an open-ended response to a question posed by the NAO, the user would touch the touchpad to signal to the NAO that it can move on to its next question. The NAO's chest light is used to signal to users the beginning (green) and end (red) of the interaction as well as when the NAO is ready to begin a new interaction phase (blue). The chest light was not used during core interaction sequences while gaze aversions were being displayed (i.e., while asking a question, answering a question, or listening to a user's utterance).

4. EVALUATION

We designed a study to test user perceptions of the generated gaze aversions and their effectiveness in enabling robots to achieve positive conversational functions in human-robot conversations. In this study, participants interacted in multiple conversational tasks with two NAO robots. Each task involved the participant either asking questions to the NAO or responding to questions posed by the NAO.

4.1. Study Design

The experiment involved a single independent variable, *gaze aversion behavior*, with three conditions varying between participants. One condition involved the robots using gaze aversions generated by the controllers described in the previous section, which we call the *good timing* condition. The other two conditions served as baselines for comparison. The first baseline was a *static gaze* condition in which the robots did not employ any gaze aversions. This baseline was included to demonstrate the importance of generating gaze aversion motions regardless of the timing. The second baseline was a *bad timing* condition in which the robot employed just as many gaze aversions as in the *good timing* condition but with reverse timings. More precisely, the *bad timing* condition produced a gaze shift toward the participant—to engage in mutual gaze—every time the *good timing* condition would have triggered a gaze aversion. Similarly, a gaze aversion is triggered every time the *good timing* condition would have produced a shift toward mutual gaze. This third condition was included as a baseline to show that not only the presence, but also the timing of gaze aversions is important for achieving positive social outcomes. Participants were randomly assigned to one of the three gaze aversion conditions, which was then held constant for all tasks (10 participants per condition). Regardless of condition, the robot always tracked the participant's face and utilized a small amount of idle head motion.

[1] http://www.aldebaran-robotics.com/

[2] http://hci.cs.wisc.edu/projects/gaze

[3] http://msdn.microsoft.com

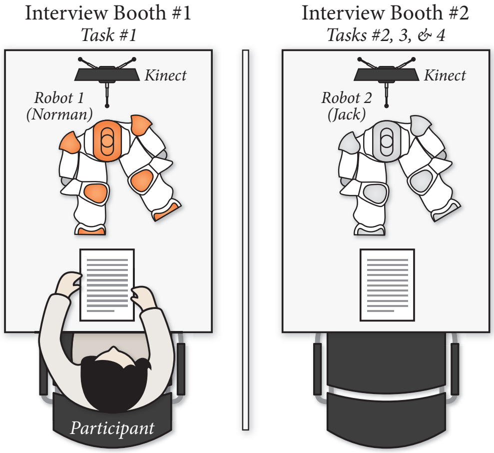

Figure 5: A diagram of the physical setup of the experiment. Participants interacted with Norman for the first task, and Jack for the other three tasks.

4.2. Participants

Thirty participants were recruited for this study (15 females and 15 males)—aged between 20 and 38 ($M = 22.90$, $SD = 4.27$)—from the University of Wisconsin–Madison campus. Participants were primarily University students with a range of fields of study, including biology, computer science, economics, and communication.

4.3. Hypotheses, Tasks, & Measures

We developed two hypotheses related to participants' perceptions of gaze aversions carried out by a robot and how robots might use gaze aversion behaviors to achieve each of the three conversational functions. The second hypothesis is split into three sub-hypotheses, one for each conversational function. Separate tasks and measures were created to test each hypothesis and sub-hypothesis. For clarity, we present each hypothesis together with the task designed to test that hypothesis as well as the primary measure in that task.

Hypothesis 1 – Well-timed robot gaze aversions will be seen as *intentional* motions used to engage in some cognitive processing, rather than randomly generated motions without meaning.

This hypothesis is derived from research that has shown how speaker gaze aversion is recognized by listeners as the speaker's attempt to disengage attention from the listener's face in order to put cognitive effort into organizing a new utterance [8, 10]. Finding support for this hypothesis enables us to conclude that the robot's head movements, which were designed to convey gaze aversion, are indeed perceived as gaze aversions and not as random movements.

Task 1 – The participants were told that the robot was training to work at a library help desk and given five library-related questions to ask the robot. They were instructed to ask each question and listen to the robot's response. The robot paused for 7 to 10 seconds (randomly determined to decrease predictability) before answering each question. Participants were instructed to ask a question again if they thought that the robot did not understand.

Measure 1 – The primary measure was the time participants waited for the robot to respond to questions before interrupting it to ask the question again. Based on our hypothesis, we expected participants to give the robot the most time to respond when it was using gaze aversions in the *good timing* condition, implying that the gaze aversion during its long pause was perceived as an intentional motion to formulate responses to the participants' questions.

Hypothesis 2 – Gaze aversions generated by our model and employed by the robot will enable it to achieve the positive conversational functions observed in human-human interaction. This hypothesis has three sub-hypotheses, one each for the cognitive, intimacy-modulating, and floor management functions of gaze aversion.

Hypothesis 2a – When a robot utilizes cognitive gaze aversions at the start of answers to questions, its answers will be rated as being more *thoughtful* and creative than when it does not display gaze aversions or displays gaze aversions with inappropriate timings.

This hypothesis is derived from research which has shown that people use gaze aversions to signal that cognitive processing is occurring and to create an impression that deep thought or creativity is being undertaken in formulating their utterance [4].

Task 2a – Participants engaged in a mock job interview with the robot, in which the participant was the interviewer and the robot was the interviewee. Participants were instructed to ask a series of five common job interview questions, and the robot was programmed to respond with answers taken from real-world job interviews.

Measure 2a – Participants rated each response immediately after it was given on four seven-point rating items. We constructed a single scale of *thoughtfulness* from the four items, including ratings on perceived thoughtfulness, creativity, disclosure, and naturalness of each response. Internal consistency was good for the items in this scale (Cronbach's $\alpha = .852$). We expected the highest overall ratings to be given by participants when the robot used cognitive gaze aversions at the start of its responses according to our model.

Hypothesis 2b – Robots that display periodic gaze aversions while listening will increase a human interlocutor's comfort and elicit more *disclosure* than robots that do not display gaze aversions or display gaze aversions with inappropriate timings.

Eye contact is one of the factors that shape feelings of intimacy between people. Too much of it results in uncomfortable levels of intimacy [4]. We posit that using gaze aversions appropriately will alleviate this potentially negative outcome.

Task 2b – The robot in this task was introduced to the participant as training to be a therapist's aide that would conduct preliminary interviews with new clients. During the task, the robot asked the participant a series of five questions with increasing levels of intimacy, ranging from "What do you like to do in your free time?" to "What is something you don't like about yourself?," and participants were instructed to respond with as much or as little detail as they wished.

Measure 2b – The primary measure for this task was the *breadth of self-disclosure*, which we obtained using a word count of participants' responses to the robot's questions. Previous research on how computers can elicit self-disclosure from people has validated word count as an appropriate measure of disclosure [17]. We expected that participants would disclose more to a robot that used appropriately timed intimacy-modulating gaze aversions while listening.

Hypothesis 2c – Robots that display gaze aversions during pauses will be perceived as *holding the floor* and will be interrupted less

Figure 6: An experimenter demonstrating the conversational interaction with Jack, one of the NAO robots.

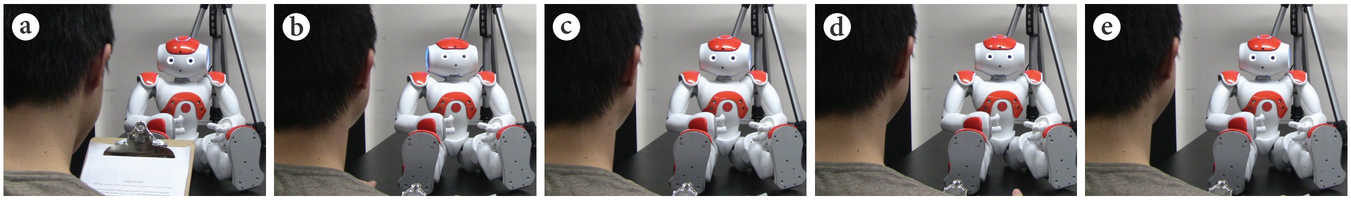

Figure 7: Example of a single question-answer sequence. (a) The participant reads a question from his list in the preparation phase. (b) The participant looks toward the robot, and the robot engages an upward cognitive gaze aversion at the start of its answer. (c) The robot looks back toward the participant during its utterance. (d) The robot engages in a sideways intimacy-modulating gaze aversion. (e) The robot looks back toward the participant to complete its utterance.

than robots that do not display gaze aversions or display gaze aversions with inappropriate timings.

Gaze has been previously shown to be important in regulating conversational turn-taking [13]. By averting its gaze at a pause in speech, we posit that a robot will be able to hold the conversational floor, whereas making eye contact during this pause will result in the robot being interrupted.

Task 2c – Participants were provided with a list of five questions to ask the robot, with the goal of getting to know each other. Participants were instructed to ask each question, listen to the robot's response, and then reciprocate with their own response to the same question. The robot's responses had two parts, separated by a pause between 2 and 4 seconds in length (randomly determined to decrease predictability). If participants started speaking during the pause, the robot refrained from giving the second part of its response.

Measure 2c – The primary measure of this task was the time participants waited for the robot to be silent during the pause in its speech before interrupting or before the robot successfully produced the second part of its utterance. We expected participants to wait longer before interrupting the robot—or to not interrupt the robot at all—when it used appropriate turn-taking gaze aversions during its speech pauses, as specified by our model.

4.4. Setup & Procedure

An orange NAO, given the name Norman, was used for the first task that tested the perceived intentionality of generated gaze shifts. A gray NAO, given the name Jack, was used for the other three tasks. A separate robot was used for the first task due to the unnaturally long pauses present in that task. We did not want negative perceptions associated with these long response times to become associated with the robot during the other three tasks in which conversations proceeded more naturally. Each robot had a unique voice, implemented as pre-recorded audio files from separate male voice actors modulated in pitch to better fit the design of the robot. Participants sat in a chair approximately three feet away from the robot they were currently interacting with. The robot's position was carefully chosen to be at approximately eye-level with the participant and at a comfortable social distance. A black dividing wall was placed between the two robots so that participants could only see a single robot at a time. The setup resembled interview booths commonly used in job fairs (Figure 5 and Figure 6).

Each question-answer interaction sequence, illustrated in Figure 7, began with a *preparation phase*—a common element of human-human interviews—in which the participant looked down and read a question from the list. During this phase, the robot displayed idle gaze movements while staying focused on the participant. Toward the end of the question, the participant redirected their attention toward the robot, at which point the robot began its response, displaying gaze aversions based on the experimental condition. Between gaze aversions, the robot displayed subtle random motions to increase lifelikeness. Upon completing its utterance, the robot always focused back on the participant, displaying subtle idle motion. At this point, the participant looked down to the list, beginning the preparation phase for the next question-answer sequence.

After obtaining informed consent, the experimenter led each participant into the study room and gave a brief introduction to the experiment. The participant first completed the single task with Norman and was then relocated to sit in front of Jack for the other three tasks, which were completed in random order. 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 received $5.

4.5. Results

We performed a mixed-design analysis of covariance (ANCOVA) to assess how robot gaze aversion behaviors affected the dependent variable for each task. Participant gender was included as a covariate to control for gender differences. Question ID—five in each task—was included as a covariate to control for learning effects. Question ID was nested within participant ID and modeled as a random effect. This resulted in five observations per participant and 150 total observations in each measure. Planned pairwise comparisons described in our hypotheses were carried out using Tukey's HSD test. A summary of our primary results is presented in Figure 8.

Hypothesis 1 predicted that gaze aversions employed by the robot would be seen as *intentional* motions used to engage in some cognitive processing, rather than randomly generated motions without meaning. Our analysis supported this hypothesis. The time given to the robot before interrupting was significantly higher when the robot used proper gaze aversion with good timing rather than bad timing, $F(1,142) = 7.72, p = .017$, or no gaze aversion at all, $F(1,142) = 5.99, p = .041$.

Hypothesis 2a predicted that when a robot utilized cognitive gaze aversions at the start of answers to participant-provided questions, those answers would be rated as being more *thoughtful* and creative than when the robot did not display gaze aversions or displayed gaze aversions with inappropriate timings. Our analysis supported this hypothesis. Participants assigned higher ratings of thoughtfulness to utterances produced by robots using well-timed gaze aversions over those produced by robots using poorly-timed gaze aversion, $F(1,142) = 27.97, p < .001$, and over those produced by robots using static gaze, $F(1,142) = 10.19, p = .005$.

Hypothesis 2b predicted that robots that displayed periodic gaze aversions while listening would increase a human interlocutor's comfort and elicit more *disclosure* than robots that did not display gaze aversions or displayed gaze aversions with inappropriate timings. Our analysis did not support this hypothesis. The robot using gaze aversions with good timing elicited no more disclosure—measured as word count per response—from participants than when its gaze aversions were badly timed, $F(1,142) = 0.56, p = .735$, or when it used no gaze aversion at all, $F(1,142) = 0.05, p = .972$.

Hypothesis 2c predicted that robots that displayed gaze aversions during pauses would be perceived as *holding the floor* and would be interrupted less than robots that did not display gaze aversions or displayed gaze aversions with inappropriate timings. Our analysis

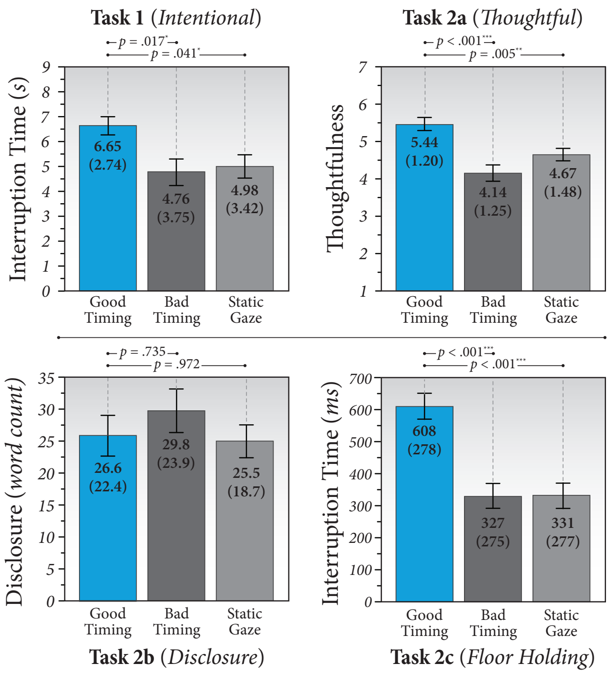

Figure 8: The results of the evaluation. Robot gaze aversions generated by our model were perceived as intentional and enabled the robot to appear more thoughtful and effectively manage the conversational floor. (*), (**), and (***) denote p < .050, p < .010, and p < .001, respectively. Means and standard deviations (in parentheses) are provided inside each bar.

supported this hypothesis. The time given to the robot during its pause before interrupting was significantly higher when the robot used properly-timed gaze aversion than when its gaze aversions were badly timed, $F(1,142) = 25.53, p < .001$, or when it did not use gaze aversion at all, $F(1,142) = 24.93, p < .001$.

5. DISCUSSION

The goal of the evaluation was to show that gaze aversions generated by our robots are perceived as intentional and meaningful motions, and that robots can use gaze aversions to achieve three conversational functions: signaling that cognitive effort is being spent in producing thoughtful utterances, modulating the overall intimacy level of the conversation, and facilitating floor management.

As shown in the first task, the robot's gaze aversions were perceived as intentional motions to engage in some sort of processing. Participants gave the robot significantly more time to formulate its response to a question when the robot used gaze aversion appropriately. Confirming this first hypothesis is important for interpreting the rest of the results, as we can conclude that the robot's head movements which were designed to convey gaze aversion were indeed perceived as gaze aversions and not as random movements.

As shown in the second task, in which a robot produced responses to a participant's interview-style questions, a robot that uses gaze aversions appropriately is capable of achieving the cognitive conversational function of producing utterances that are perceived as thoughtful and creative. This result is especially interesting as it was not observed when evaluating our previous implementation of gaze aversion behaviors for virtual agents [3]. A potential explanation might be found in the increased social presence of robots over virtual agents, making the generated cognitive gaze aversions more salient cues in this evaluation. That said, a follow-up evaluation that directly compares virtual agents and robots in this task would be required to better explain the observed difference in results.

Robots using well-timed gaze aversions were not successful in eliciting more disclosure from participants, as shown in the third task. A possible explanation is that the intimacy-modulating function of gaze aversion is really about reducing eye-contact and staring and that the precise timing for this type of gaze aversion is not important. Furthermore, the idle Perlin noise motions generated by our robot system in all conditions may have been perceived as small gaze shifts that increased participant comfort without needing more overt gaze aversions. In general, the lack of a significant result is consistent with previous HRI work that also did not find a difference in participant disclosure based on the gaze behavior of a robot [18].

Finally, robots using well-timed gaze aversion behaviors were more successful in managing conversational floor. When the robot used gaze aversions appropriately, participants waited longer during the robot's pause before interrupting it to claim the floor. It should be noted, however, that the robot was never able to hold the conversational floor throughout the entirety of its pause in speech (planned to be 2–4 seconds in length) without being interrupted. When using properly timed gaze aversions, the mean time before interruption was 608*ms*, compared with 329*ms* and 327*ms* in the *static gaze* and *bad timing* conditions, respectively. This is still a noteworthy result, as most gaps between turns in human-human conversations are shorter than 400*ms*, about one-third of them less than 200*ms* [29]. Adding other floor-holding behaviors, such as conversational fillers, might enable the robot to be even more effective in holding the floor during its pause. Previous work in HRI has already demonstrated the usefulness of conversational fillers in alleviating users' negative perceptions to long system response times [26].

5.1. Design Implications

This work has a number of design implications. In general, this work demonstrates the value of gaze aversion as a useful nonverbal conversational cue for designing applications such as an information booth robot that gives suggestions and should appear thoughtful and creative in the production of its utterances or a conversational robot that needs to effectively manage turn-taking with people. That said, gaze aversion motions alone are not sufficient for an autonomous robot system to appear natural and lifelike in its movements. In this paper, we showed how to combine aversion control with face tracking and structured random movements to create lifelike idle motion. A predictive filtering framework was very effective for gracefully combining multiple gaze motions with different goals. We have released online the code for our system of robot gaze aversion, which can be downloaded and run directly on any NAO robot with a Kinect for face tracking and speech recognition.

Another implication concerns the difference between designing social behaviors for virtual agents and physical robots. In this work, we showed how a robot could use gaze aversions effectively in conversations with people, which is similar to what we showed for virtual agents in previous work [3]. However, there are a number of important differences between virtual agents and robots that must be carefully considered when designing behaviors across both modalities, as evidenced by previous work and the difference in results across systems discussed here. It will take careful thought and creativity to develop precise representations of future agent behaviors that achieve targeted conversational functions while still being sufficiently generalizable across virtual and physical platforms.

5.2. Limitations & Future Work

One limitation of this work is that the presented gaze aversion model is static, in that the timings and frequencies of gaze aversions employed by the robot do not change over the course of the conversation or across conversations. In future work, we hope to explore how interaction history might be considered by our model. Previous research has shown that gaze aversions during speech become more common in later conversations between counselors and clients [25].

Another limitation of this work is that the gaze aversions are responsive to the speech of the human user but not to their gaze. By tracking the user's gaze, the robot might be able to more effectively regulate mutual gaze throughout the conversation and recognize gaze aversions of its user. Interactively aligning gaze aversions is particularly important to ensure that the robot's gaze aversions are recognized by the user, for instance, by producing a turn-taking gaze aversion only when the user is looking toward the robot.

When used inappropriately or in excess, gaze aversion can be perceived negatively, as illustrated by our *bad timing* condition. Previous research has linked gaze aversion with avoidance-oriented emotions, such as embarrassment, sorrow, and disgust, and direct gaze with approach-oriented emotions, such as joy, love, and anger [2]. We posit that averting gaze only when it is useful and appropriate maximizes the positive effects of eye contact. The link between gaze aversion and emotion and the tradeoff between the effects of mutual and averted gaze are important directions for future work.

Gaze aversion is only one type of gaze action, but there are many others, including fixations, glances, mutual gaze, and scanning [27]. In future work, we hope to integrate the current work on gaze aversion into a comprehensive gaze controller that implements other gaze models, and test the effectiveness of robot gaze aversion within the context of a richer set of gaze behaviors.

6. CONCLUSION

Gaze aversion—the intentional redirection away from the face of an interlocutor—is an important nonverbal cue that serves cognitive, intimacy-modulating, and floor management functions in conversations. In previous work, we identified precise spatial and temporal parameters of gaze aversion from a video corpus of human-human interactions. In this paper, we presented our work to implement these behaviors on a humanlike robot system. This implementation was designed to overcome the inherent challenges of adapting a human-based gaze model for a robot without articulated eyes. We designed a head controller for the NAO platform that generates and combines head motions with three purposes: purposeful gaze aversions to achieve conversational functions, face tracking for engaging in mutual gaze with a user, and noisy idle head motions to increase lifelikeness. An evaluation of the designed gaze aversion behaviors generated by this system demonstrated that they are perceived as intentional when expressed by a humanlike robot, and that robots can use gaze aversions to appear more thoughtful and effectively manage the conversational floor. These results have important implications for designers of human-robot interactions. Gaze aversions should be considered as an important and useful cue for developing effective conversational interactions between humans and robots.

7. ACKNOWLEDGMENTS

This research was supported by National Science Foundation awards 1017952 and 1208632 and the National Institutes of Health award 1R01HD071089-01A1. We would like to thank Faisal Khan, Brandi Hefty, Ross Luo, Brandon Smith, Catherine Steffel, and Chien-Ming Huang for their contributions to this work.

8. REFERENCES

[1] A. Abele. Functions of gaze in social interaction: Communication and monitoring. *Journal of Nonverbal Behavior*, 10(2):83–101, 1986.

[2] R. B. Adams Jr and R. E. Kleck. Effects of direct and averted gaze on the perception of facially communicated emotion. *Emotion*, 5(1):3–11, 2005.

[3] S. Andrist, B. Mutlu, and M. Gleicher. Conversational gaze aversion for virtual agents. In *Proc. IVA 2013*, pages 249–262, 2013.

[4] M. Argyle and M. Cook. *Gaze and mutual gaze*. Cambridge University Press Cambridge, 1976.

[5] M. Argyle and J. Dean. Eye-contact, distance and affiliation. *Sociometry*, 28(3):289–304, 1965.

[6] W. Bainbridge, J. Hart, E. Kim, and B. Scassellati. The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics*, 3(1):41–52, 2011.

[7] 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.

[8] G. Doherty-Sneddon and F. Phelps. Gaze aversion: A response to cognitive or social difficulty? *Memory & Cognition*, 33(4):727–733, 2005.

[9] N. Emery. The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, 24(6):581–604, 2000.

[10] 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.

[11] M. Imai, T. Kanda, T. Ono, H. Ishiguro, and K. Mase. Robot mediated round table: Analysis of the effect of robot's gaze. In *Proc. RO-MAN 2002*, pages 411–416, 2002.

[12] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, (82):35–45, 1960.

[13] A. Kendon. Some functions of gaze-direction in social interaction. *Acta psychologica*, 26(1):22–63, 1967.

[14] C. Kidd and C. Breazeal. Effect of a robot on user perceptions. In *Proc. IROS 2004*, volume 4, pages 3559–3564, 2004.

[15] K. H. Kim, R. Brent Gillespie, and B. J. Martin. Head movement control in visually guided tasks: Postural goal and optimality. *Computers in Biology and Medicine*, 37(7):1009–1019, 2007.

[16] C. Liu, C. T. Ishi, H. Ishiguro, and N. Hagita. Generation of nodding, head tilting and eye gazing for human-robot dialogue interaction. In *Proc. HRI 2012*, pages 285–292, 2012.

[17] Y. Moon. Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of Consumer Research*, 26(4):323–339, 2000.

[18] J. Mumm and B. Mutlu. Human-robot proxemics: physical and psychological distancing in human-robot interaction. In *Proc. HRI 2011*, pages 331–338, 2011.

[19] B. Mutlu, J. Forlizzi, and J. Hodgins. A storytelling robot: Modeling and evaluation of human-like gaze behavior. In *Proc. HUMANOIDS 2006*, pages 518–523, 2006.

[20] B. Mutlu, T. Kanda, J. Forlizzi, J. Hodgins, and H. Ishiguro. Conversational gaze mechanisms for humanlike robots. *ACM Transactions on Interactive Intelligent Systems*, 1(2):12:1–12:33, 2012.

[21] D. G. Novick, B. Hansen, and K. Ward. Coordinating turn-taking with gaze. In *Proc. ICSLP 1996*, pages 1888–1891, 1996.

[22] T. Pejsa, B. Mutlu, and M. Gleicher. Stylized and performative gaze for character animation. *Computer Graphics Forum*, 32(2):143–152, 2013.

[23] K. Perlin. Real time responsive animation with personality. *IEEE Transactions on Visualization and Computer Graphics*, 1(1):5–15, 1995.

[24] A. Powers, S. Kiesler, S. Fussell, and C. Torrey. Comparing a computer agent with a humanoid robot. In *Proc. HRI 2007*, pages 145–152, 2007.

[25] D. Schulman and T. Bickmore. Changes in verbal and nonverbal conversational behavior in long-term interaction. In *Proc. ICMI 2012*, pages 11–18, 2012.

[26] T. Shiwa, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita. How quickly should communication robots respond? In *Proc. HRI 2008*, pages 153–160, 2008.

[27] V. Srinivasan and R. Murphy. A survey of social gaze. In *Proc. HRI 2011*, pages 253–254, 2011.

[28] M. Staudte and M. W. Crocker. Visual attention in spoken human-robot interaction. In *Proc. HRI 2009*, pages 77–84, 2009.

[29] T. P. Wilson and D. H. Zimmerman. The structure of silence between turns in two-party conversation. *Discourse Processes*, 9(4):375–390, 1986.

[30] Y. Yoshikawa, K. Shinozawa, H. Ishiguro, N. Hagita, and T. Miyamoto. Responsive robot gaze to interaction partner. In *Proc. RSS 2006*, 2006.