A Storytelling Robot: Modeling and Evaluation of Human-like Gaze Behavior

Bilge Mutlu
Human-Computer Interaction Institute
Carnegie Mellon University
bilge@cs.cmu.edu

Jodi Forlizzi
Human-Computer Interaction Institute
and School of Design
Carnegie Mellon University
forlizzi@cs.cmu.edu

Jessica Hodgins
Robotics Institute and
Computer Science Department
Carnegie Mellon University
jkh@cs.cmu.edu

Abstract— **Engaging storytelling is a necessary skill for humanoid robots if they are to be used in education and entertainment applications. Storytelling requires that the humanoid robot be aware of its audience and able to direct its gaze in a natural way. In this paper, we explore how human gaze can be modeled and implemented on a humanoid robot to create a natural, human-like behavior for storytelling. Our gaze model integrates data collected from a human storyteller and a discourse structure model developed by Cassell and her colleagues for human-like conversational agents [1]. We used this model to direct the gaze of a humanoid robot, Honda's ASIMO, as he recited a Japanese fairy tale using a pre-recorded human voice. We assessed the efficacy of this gaze algorithm by manipulating the frequency of ASIMO's gaze between two participants and used pre and post questionnaires to assess whether participants evaluated the robot more positively and did better on a recall task when ASIMO looked at them more. We found that participants performed significantly better in recalling ASIMO's story when the robot looked at them more. Our results also showed significant differences in how men and women evaluated ASIMO based on the frequency of gaze they received from the robot. Our study adds to the growing evidence that there are many commonalities between human-human communication and human-robot communication.**

I. Introduction

Of the many applications proposed for robots with a human form, education and entertainment are two of the most promising and also likely among those with the greatest potential to be successful near term. Entertainment, in particular, can often be scripted, reducing the need for sensing and robustness to changes in the environment. Education, at least in controlled settings such as museums, has similar characteristics. Our research on humanoid robots proposes a framework of five design variables—gaze, gesture, proximity to a human partner, subtle movements of the body, speech and sound—that feature strongly in the interaction design of compelling human-robot education and entertainment applications.

In a storytelling application, all design variables will need to be scripted to act together in a natural manner. We have first chosen to focus on exploring aspects of gaze, a variable that determines how the robot should look at the members of the audience. Building on results in the literature for avatar gaze [1] and a coding of the actions of a professional storyteller, we implemented a gaze and gesture algorithm for Honda's humanoid robot ASIMO (figure 1) using a combination of hand-coded and automated procedures. The gaze algorithm was automatic, and was based on a hand-coded script of the structure of the written story. A set of generic gestures were added automatically and then supplemented by a few specialized gestures that were scripted by hand to correspond to specific content in the story.

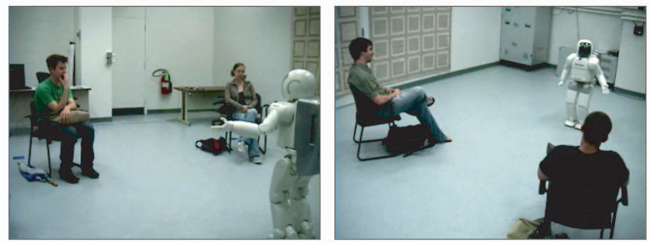

Fig. 1. ASIMO telling a Japanese fairy tale to two listeners.

Gaze is an essential component of human-human communication and is often used for communicating syntactic or semantic signals during speech [2] and storytelling in particular. Studies of gaze suggest that people who look more frequently at others are more likely to be judged favorably [3]. We used this result to assess our automatic gaze generation algorithm by manipulating the percentage of the time that the robot's gaze was directed at each of two subjects during the telling of a story. Our experimental results matched the predictions in the literature for the male subjects who, when they were gazed at less, did indeed feel less positively about the robot. To our surprise, female subjects felt more positively about the robot when they were gazed at less. Gaze is also shown to affect task performance in learning tasks [4], [5], [6]. We used this result to assess how our manipulation affected our participants' task performance. In our experiment, female subjects who were gazed at more had better recall of the story although this effect was not present for men. This experiment demonstrated that our gaze model was sufficient to reproduce the effect of at least one aspect of human gaze with a humanoid robot. It also provides further evidence that there are many commonalities between human-robot and human-human communication.

II. Background

In constructing our experimental hypotheses and design, we build on the existing literature about the social and performance-related functions of gaze in human-human communication. The following sections provide a general overview of these functions while our current research focuses primarily on oratory gaze (augmented with simple gestures) for storytelling. This section also includes a summary of existing models for implementing gaze behaviors in computer agents, avatars, and robots.

A. Social Function of Gaze Behavior

Gaze supports speech in communicating syntactic signals such as verbal utterances and emphasis [7]. Speakers direct their gaze based on the structure and content of the utterance [8]. For example, speakers look at their listener(s) less when they attempt to discuss a cognitively difficult topic [9]. We found that our professional storyteller also spent time gazing away from her listeners and included that in our model of gaze for our storytelling application.

Gaze also serves critical social functions such as communicating interpersonal attitude or affect between speaker and listener. In general, people who look more at others tend to be perceived more favorably, as more competent, friendly, credible, assertive, and socially skilled [3]. Gaze patterns also communicate liking and status among members of a group. For example, in group settings, people tend to look more at group members whom they like [9]. People look less at others of lower status [10]. Gaze patterns can communicate a speaker's attitude. Speakers tend to gaze at listeners more when they intend to be more persuasive, deceptive, ingratiating, or assertive [3].

Gender can also have a significant effect on gaze behavior. In general, women engage in eye contact more than men do [2], and are shown to look more while listening if they like the speaker. Conversely, men look more while speaking if they like the listener [9], [11].

B. Gaze and Task Performance

In addition to its social functions, gaze has significant effects on task performance. Students are shown to have significantly better recall of details of a story when their teacher made eye contact with them while reading the story than when the teacher did not [4]. Sherwood [5], in a series of five experiments, showed that students who received gaze during an oral presentation demonstrated significantly better recall than students receiving the same information without gaze. Another study showed that college students performed better in a learning task when the instructor gazed at them [6].

C. Simulating gaze behavior in agents and robots

Conversational agents have been built that model human-like gaze behavior in order to build simulations of gaze behavior in human-computer conversations [12], [1], [13]. These models include such elements of human-human communication as how speakers look away from listeners at the beginning of an utterance, and toward listeners at the end of an utterance.

Gandalf, an autonomous computer agent, used gaze to display basic attentional cues (e.g. gazing at and turning his head to the area of interest) [12]. Other applications have associated a predefined set of gaze behaviors with verbal and thematic markers [14]. For example, when the character said "Let me think..." it also looked up to indicate that cognitive processing was taking place. Garau and colleagues constructed an experiment with avatars that demonstrated that informed gaze outperformed random gaze and an audio-only condition on metrics such as how natural the conversation felt, involvement in the conversation and others [15].

Gaze has also been explored for embodied robots with the goal of creating more effective communication. Sidner and colleagues developed a head turning application that attends to users or objects in an environment and implemented it on Mel, a penguin robot [16]. They demonstrated that pointing was an effective means of communication for this robot with limited degrees of freedom. Shared attention has been a major focus on the Infanoid robot [17], [18], Robovie [19], COG [20], [21], and Kismet [22]. A number of papers have explored how to best implement gaze. For example, Imai and colleagues used Robovie to explore the accuracy required of gaze for it to be recognizable by subjects seated around a table [23]. Numerous papers have explored face tracking to facilitate gaze at particular subjects (see, for example [24], [25]).

III. Hypotheses

Drawing from these results in the social science literature, we formulated two hypotheses about responses to ASIMO's gaze behavior:

Hypothesis 1. Participants who are looked at more will perform better in the recall task than participants who are looked at less.

Hypothesis 2. Participants who are looked at more will evaluate ASIMO more positively than participants who are looked at less.

IV. Method

We designed a storytelling experience where ASIMO told a Japanese fairy tale, "The Tongue-Cut Sparrow" [26] to two listeners using a pre-recorded voice. To do so, we developed a human-like gaze model for ASIMO, creating and implementing an algorithm that dynamically directs the robot's gaze based on a coding of the story.

A. Modeling of human-like gaze behavior

Our gaze model is an extension of a model published by Cassell and colleagues [27] with parameters determined by coding the performance of a professional storyteller. Cassell and colleagues developed an empirical model of gaze behavior during turn-taking and within a turn based on the structure of the information conveyed by the speaker [27]. Their model follows the English sentence structure suggested by Halliday [28], who describes the two main structural components of an utterance using the terms "theme" and "rheme." The theme refers to the part of an utterance that sets the tone of the utterance and connects the previous utterance to the next one. The rheme contains the new information that the utterance intends to communicate. For instance, in the sentence "In the evening the old man came home." "In the evening the old man"

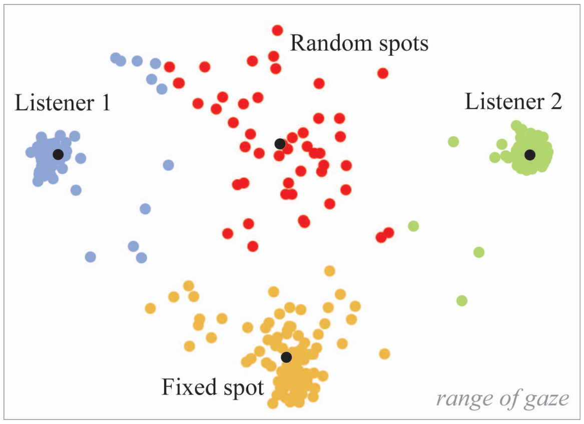

Fig. 2. Clustering of the four gaze locations used by the storyteller.

is the theme while "came home" is the rheme of the utterance. In their model, speakers look away from their listener at the beginning of a theme with 0.70 probability and look at their listeners at the beginning of a rheme with 0.73 probability. They suggested the following algorithm to simulate natural gaze behavior using a randomized function, $distribution(x)$, that returns true with probability x.

for each proposition **do**
  if proposition is theme **then**
    if beginning of turn or $distribution(0.70)$ **then**
      attach a look-away from the listener
    end if
  else if proposition is rheme **then**
    if end of turn or $distribution(0.73)$ **then**
      attach a look-toward the listener
    end if
  end if
end for

We used empirical data collected from a professional storyteller to determine locations and frequencies for this algorithm. We videotaped a professional storyteller relating two stories to a two person audience. We used 30 minutes of video data to analyze where in the environment and for how long each gaze shift executed by our storyteller was directed. Our results showed that the storyteller gazed at four different kinds of locations: the two members of the audience, a fixed spot on the table in front of her, and a set of random locations in the room. Figure 2 shows a k-means clustering of these four locations with cluster centers in black.

We defined "looking at" as keeping ASIMO's gaze on one listener once it was fixated there. "Looking away" meant looking at the other listener or looking at a random spot or the fixed location. When the gaze was not currently directed at a listener, "looking at" meant looking at one of the listeners, while "looking away" meant looking at any four of the targets with predetermined probabilities. These probabilities were derived from an analysis of the frequencies of our storyteller's gaze at each location. The duration of the gaze at each location followed a normal distribution, which we used to determine the length of the simulated gaze. Table I shows these values for each gaze location.

	Listener 1	Listener 2	Fixed spot	Random spot
Frequency (%)	13	11	38	38
Length (%)	38	27	30	5
Min (ms)	477	484	242	360
Max (ms)	15,324	5,914	13,674	4,383
Mean (ms)	2,400	2,262	2,640	1,072
Approx. StDev (ms)	500	500	500	250

TABLE I

LENGTH AND DISTRIBUTIONS OF GAZE AT EACH LOCATION.

B. Implementation

This gaze model was used with a hand-coded script of the information structure of the fairy tale to simulate human-like gaze behavior. The script marked the start of each theme and rheme and pauses between utterances. Below is the pseudo-code for our algorithm that extends the algorithm proposed by Cassell and her colleagues. In our algorithm, $distribution(x)$ produces a uniform randomized function that returns true with probability derived from [27] (e.g. 0.70) and from our empirical data. For example, $probability\_randomSpot$ is 38% from Table 1. Function $length(x)$ generates a duration for the gaze over a normal distribution with mean and standard deviation values from our empirical results ($Normal(Mean(x), StDev(x))$).

for each part of the utterance (theme/rheme/pause) **do**
  while the duration of the part **do**
    if current part is pause **then**
      if $distribution(probability\_randomSpot))$ **then**
        gaze at random spot with $length(randomSpot)$
      else
        gaze at random spot with $length(fixedSpot)$
      end if
    else if current part is theme **then**
      if $distribution(0.70)$ **then**
        if $distribution(probability\_randomSpot)$ **then**
          gaze at random spot with $length(randomSpot)$
        else
          gaze at random spot with $length(fixedSpot)$
        end if
      else
        if $distribution(probability\_listener1))$ **then**
          gaze at random spot with $length(listener1)$
        else
          gaze at random spot with $length(listener2)$
        end if
      end if
    else if current part is rheme **then**
      if $distribution(0.73)$ **then**
        if $distribution(probability\_listener1))$ **then**
          gaze at random spot with $length(listener1)$
        else
          gaze at random spot with $length(listener2)$
        end if
      else
        if $distribution(probability\_randomSpot)$ **then**
          gaze at random spot with $length(randomSpot)$
        else
          gaze at random spot with $length(fixedSpot)$
        end if
      end if
    end if
  end while
end for

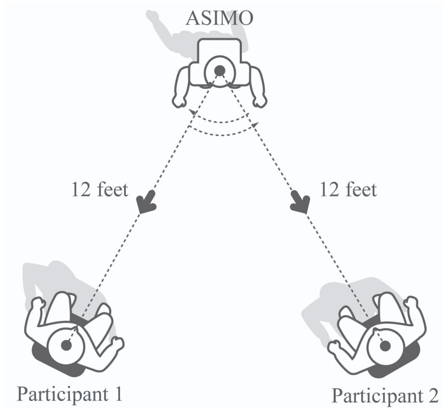

Fig. 3. Experimental setup.

The gaze algorithm was implemented on ASIMO by following a hand-coded script of the story and synchronizing ASIMO's gaze behavior with a pre-recorded voice. Ten simple arm gestures were automatically added for long utterances (greater than the mean length of 2,400 ms for gaze at a listener). Six special gestures such as bowing, crying or acting angry were added by hand when they were semantically appropriate. The location of the participants was not sensed but was determined by placing two chairs at known locations and programming ASIMO to look in those two directions. The initiation of the robot's movement was controlled by the experimenter. The robot then introduced himself to the participants, told his story, and ended the interaction.

C. Evaluation

We conducted a between-subjects experiment where participants listened to ASIMO while he told a Japanese fairy tale in English. We manipulated ASIMO's gaze behavior to gaze at one of the participants with 20% frequency and the other participant with 80% frequency. Participants were placed at the same distance from ASIMO and space was left between them so that they would not interact with each other and the robot's gaze at each participant would be easily distinguishable (Figure 3).

a) Experiment procedure: Participants were first given a brief description of the experiment procedure. After the introduction, participants were asked to answer a pre-experiment questionnaire and then provided with more detail on the task. ASIMO then introduced himself and performed the storytelling task. After listening to ASIMO's story, participants performed a distractor task, where they listened to another story on tape ("The Flying Trunk" by Hans Christian Andersen [29]). Before listening to either story, they were told that they would be asked questions regarding one of the stories. All participants were asked questions regarding ASIMO's story. After completing the task, participants answered a post-experiment questionnaire regarding their affective state, their perceptions of the robot, and their demographic information. ASIMO's story, the story on tape, and the whole experiment took an average of 17.5 minutes, 7.5 minutes, and 35 minutes respectively. The experiment was run in a dedicated space with no outside distraction. A male and a female experimenter were present in the room during the experiment. All participants were paid $10 for their participation.

b) Measures and sample: All factors in the experiment were identical for each participant except for the two controlled factors: the frequency of the robot's gaze at each participant (a manipulated independent variable) and the participant's gender (a measured independent variable). The dependent variables measured were task performance, the participant's own affective state, their positive evaluation of the robot, their perceptions of the robot's physical, social, and intellectual characteristics, their involvement in and enjoyment of the task, and participant demographics. The post-experiment questionnaire included a question as a manipulation check, "How much did the robot look at you?" Seven-point rating scales were used for all scales.

Twenty (12 males, 8 females) undergraduate and graduate students from Carnegie Mellon University participated in the experiment. Ten participants were assigned to the "looked at 80% of the time" condition. The other ten participants were assigned to the "looked at 20% of the time" condition. All participants were native English speakers and their ages ranged from 19 to 33. Participants were chosen to have a variety of majors including management sciences, social sciences, art, and engineering. Four male and three female participants had technical majors such as computer science, electrical engineering and information systems, while eight males and five females came from non-technical fields including english, business/management, writing, and psychology. On average, male participants had more video gaming experience and more familiarity with robots than female participants did.

V. Results

Our data analysis used three methods; repeated measures analysis of variance (MANOVA), regression (Least Squares Estimation), and multivariate correlations. The first method applied an Omnibus F-Test to see if the difference between pre-experiment and post-experiment measurements was significant across the two experiments, task structures, and/or genders. The second technique used a linear regression on the variables that were significant across conditions to identify the direction of main effects and interactions. The last method looked at how these variables correlated with each other. We also ran reliability tests and factor analysis on the scales we used for measurement.

Item reliabilities for all partner (robot), task, and self evaluation scales except the mutual liking scale ($\alpha = 0.54$) were high. However, since our scales for partner evaluation were created to evaluate human-like interface agents, we ran a factor analysis of all the items that we used for partner evaluation and created a highly reliable ($\alpha = 0.91$), 8-item scale for partner positive evaluation. An analysis of the manipulation check showed that the participants were aware that they were looked at more or less by the robot ($F[1:16]=3.48$, $p<0.01$).

Consistent with our first hypothesis, a regression on the performance measure showed that participants who were

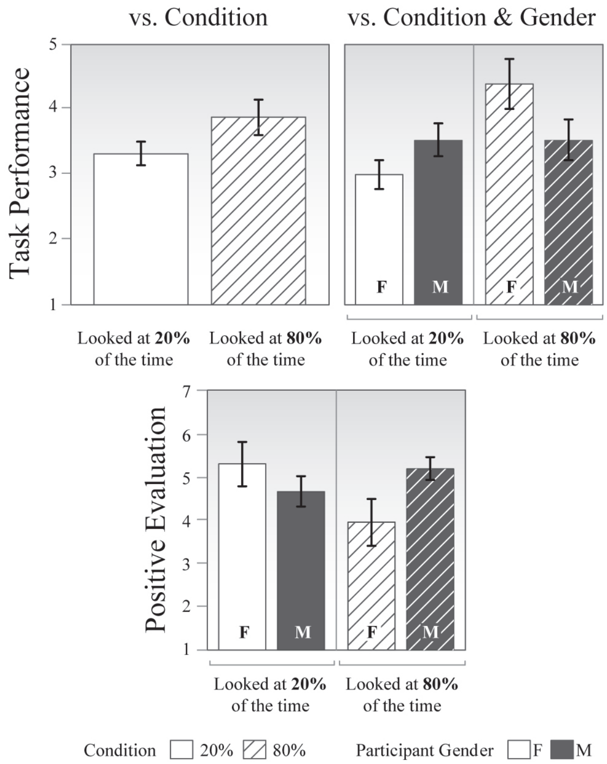

Fig. 4. *Top*: Main effect of condition and interaction between condition and participant gender on task performance. *Bottom*: Interaction between condition and participant gender on positive evaluation of the robot.

looked at more performed significantly better in the recall task (answering questions regarding ASIMO's story) than those who were looked at less ($F[1:16]=5.15$, $p=0.03$). When participant's gender was included in the statistical model, the effect was significant only in females ($F[1:16]=8.58$, $p<0.01$) while men did not show any significant difference across conditions ($F[1:16]=0$, $p=1$) (Figure 4).

Our analysis of the ratings of the positive evaluation scale showed no significant main effect but a significant interaction of experimental condition and participant gender (Figure 4). Men rated ASIMO more positively when they were looked at more while women's evaluations were higher when they were looked at less ($F[1:16]=5.62$, $p=0.03$). Although this result reveals significant interactions with participant's gender, it is not consistent with the prediction in our second hypothesis. Analysis of scales of participant's affect, task enjoyment, and task involvement did not show any significant effects or interactions.

We also looked at how our scales correlated with participant's computer use, their familiarity with robots, and video gaming experience. A multivariate analysis using Pearson's correlation coefficient showed that ratings of the positive evaluation scale was highly correlated with video gaming experience ($r=0.65$, $p<0.01$), while not correlated with computer use or familiarity with robots. This correlation held for both genders although it was stronger in men. Video gaming experience was also correlated with task enjoyment ($r=0.53$, $p=0.02$).

VI. Discussion

Our results supported the first hypothesis: the frequency of the robot's gaze affected performance on the recall task. This result has design implications for human-robot communication, particularly in education or other applications where important material is being conveyed. For example, a humanoid might try to engage a particular listener by looking at that listener more when he/she does not appear to be attending. Human-robot interactions might be designed so as to improve the recall of the material being presented.

The second hypothesis, that participants who are looked at more will evaluate the robot more positively, was not completely supported because when we included gender as a variable in that analysis, we found that women liked the robot more when they were looked at less. This result was surprising as the strong gender effect was not predicted by the literature in human gaze. However, a more comprehensive survey of results in the human-human communication literature, in particular of studies on proxemics [30], showed that this effect might be due to differences in men's and women's perceptions of personal space based on the amount of mutual gaze established with a partner [31], [32]. Bailenson et al. showed that these differences appeared in people's interactions with virtual agents [33]. They found that female participants maintained more interpersonal distance between themselves and agents who engaged them in eye contact than with agents who did not. Male participants did not show similar changes in behavior. This finding implies that because participants were not allowed to control the distance between themselves and the robot, females perhaps felt uncomfortable and evaluated the robot negatively when the robot gazed at them more. Lack of control over their distance with the robot did not affect men and they evaluated the robot more positively when the it looked at them more.

We also found that positive evaluations of ASIMO were highly correlated with participant's video gaming experience and not with their computer use, which suggests that people might perceive ASIMO as more like a video-game character or avatar than like a computer. This result suggests that we should rely most heavily on results in the interaction literature for computer agents rather than computers themselves when we design interactive experiences with humanoid robots.

Some elements of the professional storyteller's gaze were not accounted for by our model. For example, she occasionally switched from looking at one listener to looking at the other listener during a theme or rheme, but we could not find a pattern with which to model this behavior. Although we believe that our gaze model was sophisticated enough not to be perceived as algorithmic by the participants, it is possible that the introduction of more complexity based on more detailed coding of human performances would improve its naturalness. We plan to gather more data from professional storytellers and use it for the next iteration of our gaze model.

Although we were careful to make our gaze model as human-like as possible, there were still some unnatural elements in ASIMO's story telling performance. For example, ASIMO's arm gestures were found distracting by some participants, perhaps because of the servo motors that generate noise while moving the robot's arms. Another possible explanation is that our library of gestures is too limited, forcing most of the gestures to be "generic" motions of the arms to the side or front of the robot. A human storyteller would likely use gestures that were more closely matched to the content of the story. Some subjects reported that ASIMO's story was too long (17.5 minutes) and it might be easier to create a compelling performance for a shorter story.

Another limitation to the human-likeness of ASIMO's gaze model was due to the physical design of the robot. When humans direct their gaze, their movement combines movement of the eyes, the head, and the upper torso, whereas ASIMO only used head movement to shift its gaze. We used only head shift because ASIMO's design does not include controllable eyes and movement of the upper torso requires lifting and placing of the feet repeatedly, which we found to be time consuming and distracting in our pilot study.

However, our results showed that this simple head movement was sufficient to create the experimental manipulation. We asked participants to rate the amount of gaze they received from the robot. People who were looked at more thought the robot looked at them more (M=56, SD=19) and those who were looked at less thought ASIMO looked at them less (M=38, SD=20). The difference was marginally significant ($F[1:16]=3.98$, $p=0.06$). We suggest that a more sophisticated gaze model implemented on a robot with the necessary degrees of freedom would provide a more human-like gaze behavior and produce stronger social effects.

Our experiment was aided by some wizard-of-oz steps in that ASIMO did not sense where his audience was seated or allow responses from them during the telling of the story. Robust vision and natural language techniques would be required to address these issues and allow the construction of a truly interactive experience for the participants.

ACKNOWLEDGMENTS

This work was made possible by the financial support from the NSF under IIS-0121426 to the second author and ECS-0325383 and an equipment loan from Honda R&D Co., Ltd. to Takeo Kanade and the third author. We would like to thank our storyteller Gunnhildur Jonsdottir for allowing us to model her gaze.

REFERENCES

[1] J. Cassell, C. Pelachaud, N. I. Badler, M. Steedman, B. Achorn, T. Beckett, B. Douville, S. Prevost, and M. Stone, "Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents," *Computer Graphics*, 1994.

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

[3] C. L. Kleinke, "Gaze and eye contact: A research review," *Psychological Bulletin*, pp. 78–100, 1986.

[4] J. P. Otteson and C. R. Otteson, "Effects of teacher gaze on children's story recall," *Perceptual and Motor Skills*, vol. 50, pp. 35–42, 1980.

[5] J. Sherwood, "Facilitative effects of gaze upon learning." *Perceptual and Motor Skills*, vol. 2/3, no. 2, pp. 1275–1278, 1988.

[6] R. Fry and G. F. Smith, "The effects of feedback and eye contact on performance of a digit-encoding task," *Journal of Social Psychology*, vol. 96, pp. 145–146, 1975.

[7] N. Chovil, "Discourse-oriented facial displays in conversation," *Research on Language and Social Interaction*, vol. 25, pp. 163–194, 1992.

[8] A. Kendon, "Some functions of gaze direction in social interaction," *Acta Psychologica*, vol. 32, pp. 1–25, 1967.

[9] R. V. Exline and L. C. Winters, "Affect relations and mutual gaze in dyads," in *Affect, Cognition and Personality*, S. Tomkins and C. Izzard, Eds. Springer, 1965.

[10] A. Mehrabian, "Inference of attitudes from the posture, orientation, and distance of a communicator," *Journal of Consulting and Clinical Psychology*, vol. 32, pp. 296–308, 1968.

[11] A. Kendon and M. Cook, "Consistency of gaze patterns in social interaction," *British Journal of Psychology*, vol. 60, pp. 481–494, 1969.

[12] K. R. Thorisson, "Communicative humanoids: A computational model of psychosocial dialogue skills," Ph.D. dissertation, Massachusetts Institute of Technology, 1996.

[13] C. Pelachaud, N. I. Badler, and M. Steedman, "Generating facial expressions for speech," *Cognitive Science*, vol. 20, pp. 1–46, 1996.

[14] C. Pelachaud and M. Bilvi, "Modeling gaze behavior for conversational agents," in *Proceedings of the International Working Conference on Intelligent Virtual Agents*, 2003.

[15] M. Garau, M. Slater, S. Bee, and M. A. Sasse, "The impact of eye gaze on communication using humanoid avatars," in *Proceedings of ACM CHI'01*, 2001, pp. 309–316.

[16] C. Sidner, C. Lee, C. Kidd, and N. Lesh, "Explorations in engagement for humans and robots," in *IEEE Humanoids'04*, 2004.

[17] Y. Nagai, "Joint attention in infant-like robot based on head movement imitation," in *Proceedings of the International Symposium on Imitation in Animals and Artifacts*, 2005.

[18] H. Kozima, C. Nakagawa, and H. Yano, "Attention coupling as a prerequisite for social interaction," in *Proceedings of ROMAN'03*, 2003.

[19] T. Kanda, H. Ishiguro, T. Ono, M. Imai, and R. Nakatsu, "Development and evaluation of an interactive humanoid robot," in *Proceedings of IEEE IROS'02*, 2002, pp. 1848–1855.

[20] R. A. Brooks, C. Breazeal, M. Marjanovic, B. Scassellati, and M. M. Williamson, "The cog project: Building a humanoid robot," *Lecture Notes in Computer Science*, vol. 1562, pp. 52–87, 1999.

[21] B. Scassellati, "Mechanisms of shared attention for a humanoid robot," in *Embodied Cognition and Action: Papers from the 1996 AAAI Fall Symposium*. AAAI Press, 1996.

[22] C. Breazeal and B. Scassellati, "Infant-like social interactions between a robot and a human caretaker," *Adaptive Behavior*, no. 1, 2000.

[23] M. Imai, T. Kanda, T. Ono, H. Ishiguro, and K. Mase, "Robot-mediated round table: Analysis of the effect of robot's gaze," in *Proceedings of IEEE ROMAN'02*, 2002, pp. 411–416.

[24] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of CVPR'01*, 2001.

[25] M. Shiomi, T. Kanda, N. Miralles, T. Miyashita, I. Fasel, J. Movellan, and H. Ishiguro, "Face-to-face interactive humanoid robot," in *Proceedings of IEEE IROS'04*, 2004, pp. 1340–1346.

[26] Y. T. Ozaki, *The Japanese Fairy Book*. Tokyo: Tuttle Publishing, 1970.

[27] J. Cassell, O. E. Torres, and S. Prevost, *Turn Taking vs. Discourse Structure: How Best to Model Multimodal Conversation*. Kluwer, 1998.

[28] M. Halliday, *Intonation and Grammar in British English*. Mouton, 1967.

[29] H. C. Andersen, *Tales. Vol. XVII, Part 3, The Harvard Classics*. P.F. Collier & Son, 1909-14; Bartleby.com, 2001.

[30] E. Hall, "System for the notation of proxemic behavior," *American Anthropologist*, vol. 65, pp. 1003–1026, 1963.

[31] M. Argyle and J. Dean, "Eye-contact, distance and affiliation," *Sociometry*, vol. 28, pp. 289–304, 1965.

[32] L. A. Hayduk, "Personal space: Where we now stand." *Psychological Bulletin*, vol. 94, pp. 293–335, 1983.

[33] J. Bailenson, J. Blascovich, A. Beall, and J. Loomis, "Equilibrium revisited: Mutual gaze and personal space in virtual environments," *PRESENCE: Teleoperators and Virtual Environments*, vol. 10, pp. 583–598, 2001.