Designing Effective Gaze Mechanisms for Virtual Agents

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

ABSTRACT
Virtual agents hold great promise in human-computer interaction with their ability to afford embodied interaction using nonverbal human communicative cues. Gaze cues are particularly important to achieve significant high-level outcomes such as improved learning and feelings of rapport. Our goal is to explore how agents might achieve such outcomes through seemingly subtle changes in gaze behavior and what design variables for gaze might lead to such positive outcomes. Drawing on research in human physiology, we developed a model of gaze behavior to capture these key design variables. In a user study, we investigated how manipulations in these variables might improve affiliation with the agent and learning. The results showed that an agent using affiliative gaze elicited more positive feelings of connection, while an agent using referential gaze improved participants' learning. Our model and findings offer guidelines for the design of effective gaze behaviors for virtual agents.

Author Keywords
Nonverbal behavior; gaze; virtual agents; affiliation; learning

ACM Classification Keywords
H.1.2 Models and Principles: User/Machine Systems—*Human factors*; H.5.2 Information Interfaces and Presentation: User Interfaces—*Evaluation/methodology, User-centered design*

General Terms
Design, Experimentation, Human Factors

INTRODUCTION
Virtual agents hold tremendous potential for computer interfaces with their unique ability to embody humanlike attributes [7]. These attributes form rich communication mechanisms with which people are intimately familiar and afford intuitive interactions. Variations in these attributes activate key social and cognitive processes, in turn eliciting significant positive outcomes such as improved learning and rapport in key application domains such as education [33], collaboration [46], and therapy [49]. A deeper understanding of how humans use these attributes in social interactions might enable designers to create more effective virtual agents in these domains.

Figure 1. Erin, one of the humanlike virtual agents we used in our study which examined how virtual agents could use their gaze effectively in an educational scenario. Here Erin is giving a lecture on geographical locations of ancient China.

Gaze cues are an important subset of the embodied cues that people employ in social interactions. Using gaze cues, people can control the flow of a conversation, indicate interest in or appraisal of objects and people, improve listeners' comprehension, express complex emotions, and facilitate interpersonal processes [1, 3, 36]. In order to design virtual agents that achieve such significant effects, we must first design mechanisms that capture key variables of gaze and generate appropriate gaze behaviors. However, creating effective gaze cues for virtual characters is an open challenge, as humans have built and finessed subtle but complex patterns in which they use these cues over thousands of years of evolution and developed a sensitivity to observing them in others [42]. Furthermore, how these design variables might be varied to achieve specific high-level social and cognitive outcomes is not well understood. The goal of our research is to explore how agents might achieve these outcomes through subtle variations in gaze behaviors and what design variables might enable such variations. We contextualize the work presented in this paper in an educational scenario, in which virtual agents promise improvements in affiliation with students and in learning.

In this paper, we demonstrate that virtual agents can use subtle variations in gaze to achieve significant high-level outcomes. We present a parametric model that synthesizes appropriate gaze behaviors and allows for systematic variations in these behaviors, which we validated through an online user study with 96 participants. Our design of the gaze model draws on research in human physiology, which precisely describes

CHI'12, May 5–10, 2012, Austin, Texas, USA.

how humans coordinate various gaze parameters such as the movements of the eyes and the head when executing *gaze shifts*—the intentional redirection of gaze toward a particular piece of information in the context of interaction. Gaze shifts are a fundamental building block of overall gaze behavior. Through subtle variation in timing and movement of the head and eyes, individuals construct a range of high-level behaviors. Our model of gaze was designed and parameterized in such a way as to provide sufficient control over the subtleties of head and eye movements, enabling a virtual agent to naturally display these behaviors.

We also present a user study with 20 participants, in which we investigated how a virtual agent (Figure 1) might manipulate low-level gaze parameters to achieve high-level effects such as improved learning and feelings of rapport in a storytelling context. We manipulated the parameters of our gaze model to create *affiliative gaze*—maintaining a head orientation toward the participant to emphasize the social interaction [12]—and *referential gaze*—maintaining a head orientation toward shared visual space to emphasize the information to which the speaker is referring [30]. In both cases, the agent looks at the same targets; the difference lies in the timing and the degree to which the agent uses its head or eyes to look at these targets. The results showed that these manipulations generated significant social and cognitive effects; the use of affiliative gaze improved the perceptions of the agent and the use of referential gaze increased recall performance.

The remainder of the paper provides related work on gaze from a number of perspectives, describes our model and a validation study, outlines the design and results of our user study, and discusses our findings and their implications for the design of effective gaze mechanisms for virtual agents.

BACKGROUND

Gaze is an important social cue that has been studied under a number of different academic disciplines. First we present work in psychology on how humans use their gaze in social interactions, as well as some of the achievable high-level effects of gaze. Next we discuss related work on embodied conversational agents in the computer graphics and HCI literatures.

Gaze in Human Communication

In general, being gazed at by another person can result in a number of interesting effects. Gaze is predominantly interpreted as being attended to, and depending on the context of interaction, being gazed at can lead to discomfort from feeling observed, or lead to genuine social interaction [1]. A person who makes increased eye contact is associated with greater perceived dynamism, likeability, and believability [4]. These people are also seen as more truthful or credible [1]. One illustrative study specifically showed that people who spend more time gazing at an interviewer receive higher socioemotional evaluations [16]. On the other hand, gaze aversion produces consistently negative effects in impressions of attraction, credibility, and relational communication [5].

An important construct in the study of nonverbal behavior is *immediacy*, defined as the degree of perceived physical or psychological closeness between people [37]. Gaze has been found to be a significant component of immediacy, especially in the context of improving educational outcomes [21]. Students from primary school age through college have been shown to learn better when they are gazed at by the lecturer [41, 47]. Learning in these cases was usually measured by the students' performance on recall tasks. In a similar study, gazing into the camera during a video link conversation was shown to increase the recall of the viewer at the other end [14]. Gaze's positive effect on recall is usually attributed to its role as an arousal stimulus, which increases attentional focus and therefore enhances memory [26].

The above work shows that gaze as a whole can create positive interactions and learning effects. Our goal is to investigate how subtle parameters of gaze might serve to strengthen or weaken these effects. One of these parameters is the alignment of the head, which plays a substantial role in shifting a viewer's visual attention [23]. Gazing at someone with the head fully aligned (affiliative gaze) might be perceived differently than gazing at someone out of the corner of the eyes with the head aligned towards information of interest in the context of interaction (referential gaze). A virtual agent might be able to manipulate this, and other parameters, to achieve specifically desired effects.

Embodied Conversational Agents

Our work builds on previous research in the area of embodied conversational agents. Developing embodied agents that can communicate effectively is a significant area of research. Giving these agents more complex and human-like behaviors has been a longstanding goal, for example to make inhabited interfaces and agents in VR environments more effective [40]. The ability to use non-verbal communicative behavior is very important as it increases positive feelings of copresence and familiarity, and in general makes agents more effective communicators [2].

An agent's gaze behavior is very important to providing rich interactions; well-designed gaze mechanisms–e.g., gazing at turn-taking boundaries during conversation–can result in more efficient task performance and more positive subjective evaluations [22]. However, poor gaze behavior can be worse than no gaze at all. The positive effects of having an embodied agent–as opposed to only audio or text–can be completely lost if the gaze is very poor or random [15]. Gaze behavior of a female virtual agent, when coupled with the agent's appearance, can make the difference between enhancing negative attitudes toward women or breaking gender stereotypes [10]. The high-level effects of gaze discussed above have been shown to also extend to physical embodiments, such as robots. For example, the learning effect of gaze has been observed in human-robot interactions; increased gaze from a storytelling robot facilitates greater recall of the story [39].

Previous work on modeling human gaze to inform the design of agents includes the area of turn management [8, 43], figuring out where agents should be looking and why they should be looking there [25, 31], or how an agent should make random eye saccades in idle situations [6] or face-to-face conversations [32]. Some attempts have been made to model head and eye motions for virtual agent gaze, including data-driven [9]

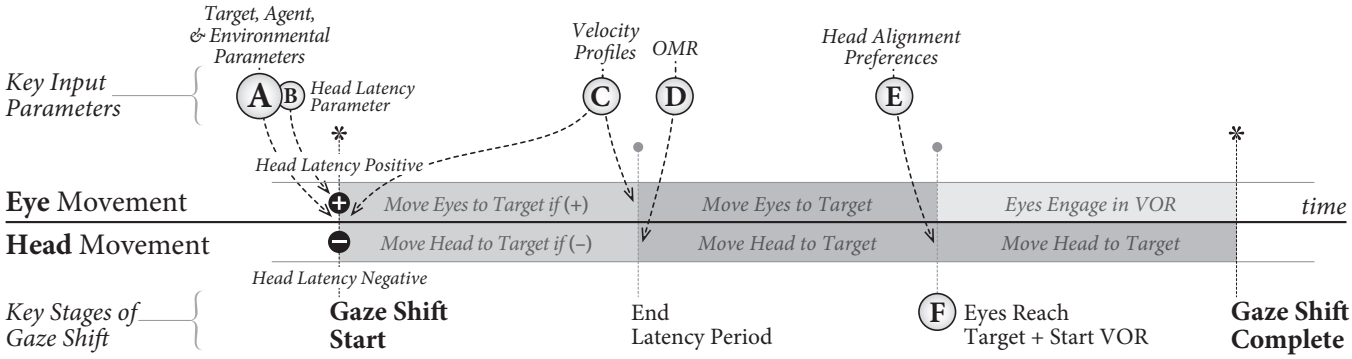

Figure 2. A visual representation of our model of gaze shifts mapped onto time. Key input variables and processes include (A) target, agent, and environmental parameters, (B) head latency, (C) velocity profiles for head and eye motion, (D) oculomotor range (OMR) specifications, (E) head alignment preferences, and (F) the vestibulo-ocular reflex (VOR)[1].

and procedural [45] approaches, as well as a hybrid between the two [29]. In general, existing research has proven that more realistic gaze behavior for humanoid avatars results in considerably improved and fluid communication [15]. At the very least, simply *conveying* gaze direction eases turntaking and is essential to establishing who is talking and listening to whom in multiparty mediated communication [50].

While previous research has explored how gross manipulations in gaze behavior might shape user experience and perceptions of virtual characters, how specific parameters in the control space for gaze might be mapped to specific outcomes in interaction has not been explored. Our work seeks to fill this knowledge gap. Our model of gaze, presented next, provides parameters that work as low-level gaze variables, and allows us to investigate their impact on high-level outcomes in a principled way. Advantages of our model include its procedural simplicity, grounded nature, and proven ability to produce large objective and subjective high-level effects through the manipulation of very subtle parameters.

DESIGNING EFFECTIVE GAZE MECHANISMS

In order to design virtual agents to be effective communicators, we must develop a deeper understanding of how humans use subtle gaze behavior to achieve high-level outcomes. We turn to research in neurophysiology to develop this understanding. Neurophysiologists have studied how humans and other primates coordinate head and eye movements during gaze shifts, in the process revealing potential parameters of gaze that might eventually lead to the positive social outcomes we wish to create with virtual agents. Here we present the most relevant findings from neurophysiology literature which we combine into a model of gaze shifts useful to virtual agents. We also present a validation of our model, taking the form of a user study, which shows that virtual agents using the model can execute gaze shifts that are both communicative and natural. [1]

Given the current configuration of the head and eyes and input parameters indicating movement characteristics and target gaze direction, our model computes a trajectory for the head and eyes. It first computes a few key variables for the movement, and then computes velocities for the head and eye rotations for each frame of the animated movement, keeping the gaze on target until the eyes and head have reached their final target rotations. The model is presented graphically in Figure 2, and includes six main components: (A) target, agent, and environmental parameters, (B) head latency, (C) velocity profiles for head and eye motion, (D) oculomotor range (OMR) specifications, (E) head alignment preferences, and (F) the vestibulo-ocular reflex (VOR).

The first variable to compute when executing a gaze shift is head movement latency in relationship to the eye movement (Figure 2b). This head latency can vary from person to person and task to task. Factors such as target amplitude, the predictability of the target [44], modality of the target (visual or auditory) [17, 18], target saliency, vigilance of the subject, and whether the gaze shift is forced or natural [51] have an effect. These factors serve as the target, agent, and environmental input parameters to our model (Figure 2a). Our model stochastically combines these parameters to determine the head latency based on findings from Zangemeister et al. [51]. A positive latency results in a period of eye motion during which the head remains fixed, while a negative latency result in a period of head motion while the eyes remain fixed.

Once the head latency period is complete, both the eyes and head begin simultaneously moving towards the target. Both during and after the latency period, each eye and the head follow velocity profiles (Figure 2c) that resemble standard ease-in and ease-out functions [27, 32]. These movements prevent unnatural head motions caused by high-frequency signals [35]. The maximum velocity of the eyes and head—the peaks of the velocity profiles—are computed based on positive linear relationships with the amplitude of the intended gaze shift [20]. Once the maximum velocity for the gaze shift is computed, the shift is executed by computing the actual velocity for each frame of the animation. The velocity is determined by a piecewise polynomial function derived to approximate the published experimental data. This polynomial function can be expressed as follows, where g is the proportion of the gaze shift completed, V_{max} is the maximum velocity, and V is the current calculated velocity.

[1] A full specification of the model, including documented pseudocode, full details of the validation, and videos of animated characters using the model for their own gaze behavior, can be found at http://hci.cs.wisc.edu/projects/gaze.

$$V = \begin{cases} 2V_{max} \cdot g & g \in [0, 0.5) \\ 4V_{max} \cdot g^2 - 8V_{max} \cdot g + 4V_{max} & g \in [0.5, 1] \end{cases}$$

An important component of the model is the oculomotor range (OMR) (Figure 2d), which limits rotation of the eyes such that they don't roll back into the head. The human OMR has been estimated to be between $45°$ and $55°$. A virtual character's baseline OMR can be determined based on the size of the eye cavities and the size of its pupils and irises. However, merely encoding these OMR values as static parameters is not sufficient, as the *effective* OMR may fluctuate during the course of a single gaze shift. The fluctuation is a product of the neural (as opposed to mechanical) nature of the limitation imposed on eye motion. The virtual agent's eye rotations are never allowed to surpass the current *effective* OMR.

At the onset of a gaze shift, effective OMR is computed based on the initial eye position and the baseline OMR. Initial eye position is measured in degrees as a rotational offset of the current eye orientation from a central (in-head) orientation. This value is only non-zero when the rotational offset is contralateral (on the opposite side of center) to the target. When the eyes begin the gaze shift at these angles, the effective OMR has a value close to the original baseline value. When the eyes begin the gaze shift closer to a central orientation in the head, the effective OMR diminishes [11]. Effective OMR is also updated throughout the gaze shift at every time step according to the concurrent head velocity. As the head moves faster, the effective OMR diminishes [20]. See the online supplement for details on the effective OMR computations, including equations and empirically derived constants.

The next component of the model is a user-defined parameter specifying the *head alignment* preferences of the agent (Figure 2e). Head alignment—how much individuals use their heads in performing a gaze shift—is highly idiosyncratic, and creates a differential directness of gaze at the target [13]. A parameter value of 0% for head alignment indicates that once the eyes have reached the gaze target, the head stops moving, resulting in gaze at the target out of the corner of the eyes. On the other hand, at a 100% parameter value for head alignment, the head continues rotating until it is completely aligned with the target, with concomitant compensatory eye movement to keep the eyes directed toward the target, resulting in gaze with the eyes and head both fully directed towards the target. Head alignment values between these two extremes can be computed using spherical linear interpolation between the two corresponding rotational values. This parameter can be manipulated to create affiliative gaze–keeping high head alignment with a conversational partner and low head alignment with everything else–or referential gaze–keeping high head alignment with information being referred to in the environment and low head alignment with the conversational partner. Head alignment is the most directly controllable input parameter to the model, and as such is the parameter we focus on in our experimental methodology.

Once the eyes have reached their target, the vestibulo-ocular reflex (VOR) keeps them locked to the target while the head finishes its rotation (Figure 2f). This component of gaze, the final in our model, is handled by rotating the eyes in the opposite direction of head motion as the head completes its portion of the gaze shift, keeping the eyes fixated on the gaze target even as the head keeps rotating.

Ancillary components of our model include a blink controller for generating both random and gaze-evoked blinking as described by Peters [45], and idle gaze behavior. When the agent is not actively engaging in a gaze shift following our model, the eyes are controlled by an implementation of the model presented in Lee et al. [32]. This implementation of subtle random eye movements has been shown to dramatically increase the realism of the character. The virtual character's eyelids also move with vertical shifts of the eyes, rising up as the eyes pitch upwards and dropping down as the eyes pitch downward [48]. Lastly, we implemented ambient facial and bodily cues for our virtual characters to prevent the unnatural rigidity inherent to a static character. When the character is in an idle state, these cues include smiles and slight movements of the arms and the head.

Model Validation

Our model development was followed by an empirical evaluation of the communicative accuracy and perceived naturalness of gaze shifts generated by our model. We compared gaze shifts generated by our model against those generated by a baseline model and those displayed by a human.

Participants – Ninety-six participants (50 males and 46 females) took part in the validation study. The participants were recruited through Amazon.com's Mechanical Turk online marketplace, following crowd-sourcing best practices to minimize the risk of abuse and to achieve a wide range of demographic representation [28, 24]. Participants received $2.50.

Study Design – In the study, participants were shown a series of videos in which either an animated virtual character or a human confederate shifted gaze toward one of sixteen objects arranged on a desk. This simplified scenario allowed us to focus our evaluation on the effectiveness and naturalness of gaze shifts, while minimizing contextual and interactional

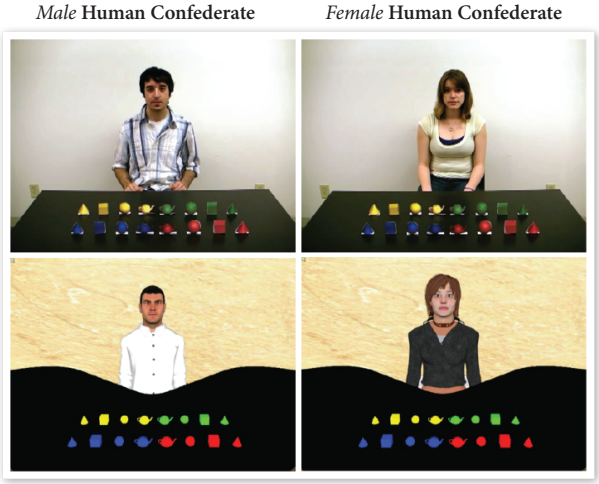

Figure 3. Still images from the videos presented to the participants.

factors and facilitating the matching of animated and real world conditions. Participants observed the agent from the perspective of a collaborator seated across from the agent or the human confederate. The objects on the desk were separated into four groups, distinguished by color and shape. Still images from the videos are shown in Figure 3.

The study followed a two-by-four split-plot design. The factors were gender–participant matched with agent–varying between participants, and model type, varying within participants. The model type independent variable included gaze shifts generated by a baseline gaze model from Peters [45], and those produced by our model. The model type independent variable also included two control conditions. In the first control condition, a virtual agent with a male or female appearance maintained gaze toward the participant (i.e., the camera) without producing any gaze shifts at all. In the second control condition, a male or female human confederate presented gaze shifts toward the object on a desk in front of him/her. The order in which the conditions were presented to each participant was counterbalanced.

Procedure – Each participant was shown 32 videos of a virtual character or human. In the videos, the agents or the confederates gazed toward the participant (i.e., the camera), announced that they are about to look toward an object of a specific color on the table, shifted their gaze toward the object, and moved their gaze back toward the participant. Following each video, the participants answered a set of questions that measured the dependent variables. Participants evaluated each condition four times in a stratified order. Each video was approximately 10 seconds, with the overall study lasting around 20 minutes.

Measures – The validation study used two dependent variables: *communicative accuracy* and *perceived naturalness*. Communicative accuracy was measured by capturing whether participants correctly identified the object toward which the gaze shift of the agent was directed. Perceived naturalness was measured with a four-item scale with high reliability, including *naturalness*, *humanlikeness*, *lifelikeness*, and *realism*, Cronbach's $\alpha = .94$.

Results – We conducted a mixed-model analysis of variance (ANOVA) on the data to determine the effect that different gaze models had on how accurately participants identified the object that the agents or the human confederates looked toward and the perceived naturalness of the gaze shifts. Overall, model type had a significant effect on communicative accuracy, $F(7, 46.67) = 16.77, p < .001$, and perceived naturalness, $F(7, 46.67) = 151.03, p < .001$.

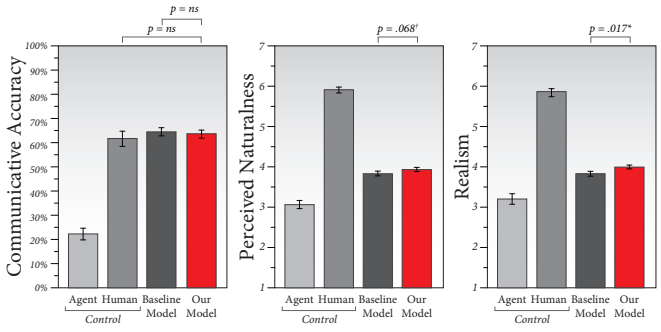

Figure 4. Results from the communicative accuracy, perceived naturalness, and realism measures. Conditions are from left-to-right: agent control (no gaze shifts, only idle gaze), human control (video of a gazing human), agents shifting their gaze using a previously published baseline gaze model [45], and agents shifting their gaze using our model of gaze.

Pairwise comparisons found no significant differences in the *communicative accuracy* of the gaze shifts produced by our model and those produced by human confederates, $F(1, 14.91) = 0.03, p = ns$. Similarly, no differences in accuracy were found between our model and the baseline model, $F(1, 1958) = 0.17, p = ns$. The results suggest that the gaze shifts generated by our model are just as accurate as those performed by human confederates and those generated by the baseline model.

Comparisons across conditions showed that participants rated gaze shifts generated by our model as marginally more *natural* than those generated by the baseline model, $F(1, 1967) = 3.34, p = .068$. Comparisons over *realism* (one of the items included in the perceived naturalness scale) found that gaze shifts produced by our model were rated as significantly more realistic than those generated by the baseline model, $F(1, 1963) = 5.75, p = .017$. Results on the communicative accuracy, perceived naturalness, and realism measures are illustrated in Figure 4.

EXPERIMENTAL EVALUATION

The validated model of gaze shifts provides parameters that allow us to explore how manipulation of these low-level parameters can achieve valuable high-level effects. The main study of this paper explores manipulating the *head alignment* parameter in order to create more affiliative or more referential gaze cues, with the goal of increasing feelings of connection with the agent and increasing learning respectively. The experiment was carried out in the context of an educational scenario, with the virtual agent serving as a lecturer to the human participant. The virtual agent taught the participant about a specific subject from ancient Chinese history. A map of China was used as a reference to facilitate the descriptions of geographical locations. An example of the interface presented to participants is shown in Figure 1.

Hypotheses

We designed an experiment to test three hypotheses. First, we wished to merely confirm that the presence of an agent will elicit better learning than only audio.

Hypothesis 1 – The presence of an embodied agent will result in better recall performance than only hearing audio with no accompanying agent.

The literature strongly supports the first hypothesis, in which it is shown that teachers who gaze at their students lead the students to learn better [41, 47]. We seek to show these same effects can be achieved by virtual agents using our gaze model.

Hypothesis 2 – An agent which employs more affiliative gaze (maintains higher head alignment with the participant) will garner higher subjective evaluations than one which uses more

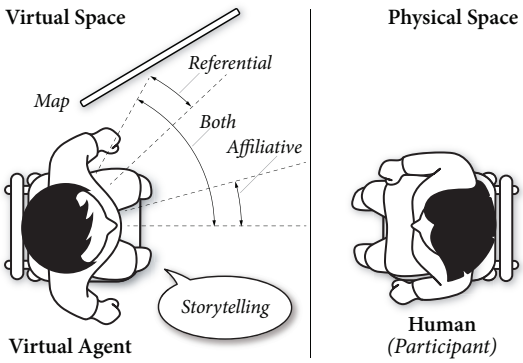

Figure 5. A diagram of the setup of the study showing the range of the agent's head movements for each gaze condition. The agent's eye motions (not depicted) always move the full distance from eye contact with the participant to eye contact with each map location being referred to.

referential gaze (maintains higher head alignment with the information being referred to).

This hypothesis is supported by the fact that people are perceived as being more intelligent, more trustworthy, and more friendly if they make more direct eye contact [1, 14]. We wish to extend this work to show that head alignment makes an impact on the positive subjective effects of mutual gaze, and that our model is capable of achieving these effects.

Hypothesis 3 – Viewing an agent using more referential gaze should result in better recall performance. This will especially be true when the information to be recalled relies on building associations to objects in the environment.

Referential gaze helps build associations between information and objects in the environment. In communication between a human and virtual agent with a shared environment, being able to perceive the agent's eye gaze to different objects in the environment reduces the time and verbal communication needed for grounding references [34]. We believe that head alignment has the ability to strengthen or weaken this effect since it plays a substantial role in shifting a viewer's visual attention [23]. When the agent's head is aligned more fully with the object under consideration, this should serve as a stronger referential gaze cue than if the head is not aligned. We wish to show that our gaze model can achieve this effect.

Participants

We recruited 20 participants for our study (10 males and 10 females), with ages ranging from 19 to 65 ($M = 27.8$, $SD = 14.3$). Fifteen participants were students and five were working members of the community. All were native English speakers. Student participants came from a number of different fields, including psychology, engineering, and business. All participants were recruited using a combination of campus flyers and on-line student job forums.

Study Design

We employed a within-subjects design for this study. Our experiment involved one factor, *type of gaze*, with four levels, each of which is defined as follows:

- **Audio**: The agent is shown briefly during its introduction. Then the lights in the virtual scene are extinguished for the duration of the lecture, except for a spotlight on the map, so only the map can be seen and the agent is in complete darkness. At the end of the lecture, the scene lights turn back on for the agent to give its instructions to the participant on taking the subjective evaluation and quiz for the lecture.

- **Affiliative**: The agent keeps its head aligned with the participant as much as possible during the lecture. When the agent is making direct eye contact with the participant, the head is fully aligned with the participant. When the agent shifts its eye gaze to refer to something on the map, the head aligns as little as possible with the map – as much as the agent's OMR will allow – so as to keep the head aligned towards the participant.

- **Referential**: The agent keeps its head aligned with the map as much as possible during the lecture. When the agent is gazing at the map, the head is fully aligned with the map. When the agent shifts its eye gaze back to the participant, the head aligns as little as possible with the participant – as much as the agent's OMR will allow – so as to keep the head aligned towards the map. The referential condition is "more referential" (borrowing the term "referential" from linguistics) because the head maintains alignment with the information being referred to on the map, both when looking at the participant (out of the corner of the eyes) and when looking at the map (head and eyes fully aligned).

- **Both**: When gazing at the participant, the agent keeps its head fully aligned to the participant. When gazing at the map, the agent aligns its head fully with the map.

Figure 5 and Figure 6 illustrate the difference in head motion between the three conditions *affiliative*, *referential*, and *both*. It should be stressed that in all three of these conditions, the agent's eyes move the full distance from eye contact with the participant to each map location being referred to. The agent's eyes always converged on each map location while it was being referred to.

Each participant viewed four lectures given by four different virtual agents, each utilizing a different gaze condition. Thus, every participant was exposed to all four gaze conditions. All agents, three of which can be seen in Figures 1 and 6, were specifically designed to look and sound androgynous so as to eliminate gender biases as much as possible. Each agent always gave the same lecture, but pairings between agent and gaze condition were stratified across participants for balance.

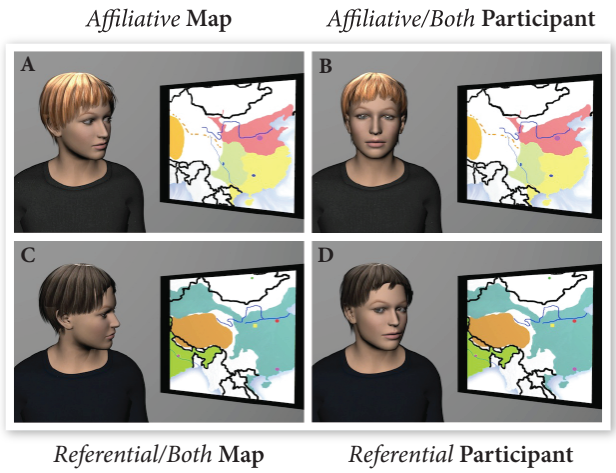

Figure 6. A visual depiction of an agent in different gaze conditions: (A) Affiliative, looking at map, (B) Affiliative (and Both), looking at participant, (C) Referential (and Both), looking at map, (D) Referential, looking at participant.

Figure 7. The physical setup of the experiment.

The order in which the lectures – and accordingly gaze conditions – were presented to the participants was completely randomized to offset any bias due to fatigue. Lectures were also kept informationally distinct in an attempt to decrease learning bias. Each lecture was carefully controlled to be near the same length–approximately three minutes.

During each lecture, the agent makes eleven gaze shifts to the map, each followed shortly by a gaze shift back to the participant. The agent fixates its gaze on different targets on the map as they are mentioned, e.g. while giving facts about a Chinese city visible on the map. Timing of this gaze shift is done according to Griffin [19] and Meyer et al. [38], which indicate that a deictic gaze shift should occur 800 - 1000 ms before the object being gazed at is mentioned in speech.

Implementation – The experiment was implemented using a custom framework built on top of the Unity game engine (www.unity3d.com). The gaze model was implemented as a Unity script. The characters were created using a commercially available parametric base figure. Audio was prerecorded and pitch-shifted to sound androgynous.

Procedure

The experiment was conducted in a closed study room with no outside distraction. Participants entered the experiment room and were asked to sit at a table with a single computer monitor and mouse. The monitor was a 32-inch flat panel display, allowing the virtual agent representation (only head and shoulders) to be near life size. This setup can be seen in Figure 7. The experimenter gave the participant a brief description of what they would be asked to do in the experiment, and then asked them to review and sign a consent form. The experimenter told the participants that they were going to be listening to and quizzed on four short lectures from different virtual lecturers, each on a topic pertaining to ancient China.

After consenting, the experimenter told participants that they could press the start button on the screen after he left the room. Upon pressing the start button, the first lecturer (randomly chosen by the software) began introducing itself and giving its lecture. Upon completion of the lecture the screen went black, and participants filled out on paper a subjective evaluation of the lecturer they had just viewed, followed by a quiz on the material presented in the lecture. During the initial instructions, the experimenter made it clear that the quiz was to be filled out *after* the subjective evaluation had been completed. This allows the subjective evaluation to double as distractor task, strengthening any subsequent recall measures. All participants were monitored via closed-circuit camera by the experimenter to ensure that these instructions were followed, but no participants were observed to neglect the instructions.

After filling out the subjective evaluation and quiz, the participant could go back to the monitor and click the on-screen button to begin the next lecture. This process was repeated until all four lectures had been viewed, rated, and quizzed. At this point, the experimenter re-entered the room with a short questionnaire of demographic information. Following completion of the questionnaire, the experimenter debriefed and payed the participant. The total experiment took approximately 30 minutes, and participants were paid $5.

Measures

Our experiment involved one independent manipulated variable, *type of gaze*, manipulated within participants. The dependent variables included objective measurements for evaluating participants' recall of the lecture material and subjective measurements for evaluating the participants' impressions of the virtual agent.

The objective measurement of recall involved quizzes taken by all participants following each lecture. Each quiz had ten short-answer questions. These questions were split into three categories (not visible to the participant). One category included three questions that asked about information not directly associated with information on the map. Thus, referential gaze should not have had an effect on the recall of this information. An example question in this category would be, "In what year did the Jin dynasty overtake control of China?" The Jin dynasty was not represented on the map during the associated lecture. The second category, consisting of four questions, asked about purely spatial information. For example, one question here was "Which of the *Three Kingdoms* dynasties extended farthest south?" This question is only answerable by having studied the map. The third category, including the remaining three questions of the quiz, relied on building associations between verbal lecture content and spatial map locations. For example, "Give one reason why Emperor Wen declared the city of Luoyang to be his capital city." Referential gaze was expected to make the biggest impact on questions from the latter two categories. Questions from all three categories were randomly permutated to create the final ten-question quiz.

The subjective measurements were split into six broad indicators. Each question within the indicators took the form of a seven-point rating scale. Item reliability (Cronbach's α) was acceptable or better for all except *skilled communicator*.

1. *Likeability*: Four-item measure of how likeable the participant found the agent to be. Includes questions on perceived friendliness and helpfulness. (Cronbach's $\alpha = .78$)

2. *Rapport*: Six-item measure of how much the participant felt feelings of rapport in relation to the agent. Questions asked, e.g., how well the participant felt he or she connected with the agent and how willingly he or she would disclose personal information to the virtual agent following the lecture. (Cronbach's $\alpha = .84$)

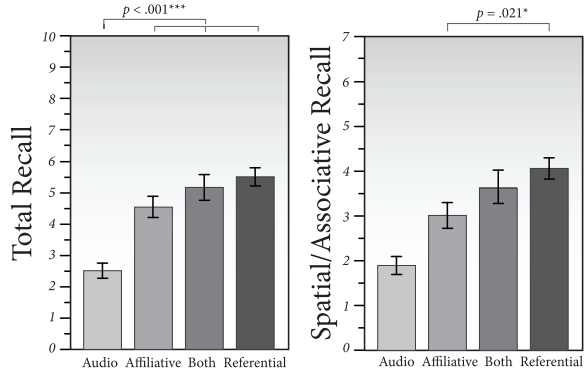

Figure 9. Objective measure (recall measured by post-lecture quiz). On the left is the total quiz performance, on the right is the quiz performance when only considering a subset of the questions: those dealing with spatial information and building associations.

3. *Trust*: Two-item measure of how trustworthy the participant perceived the agent to be. Includes ratings of trustworthiness and honesty. (Cronbach's $\alpha = .72$)

4. *Intelligence*: Three-item measure of how intelligent the participant perceived the agent to be. Includes ratings of competence and expertise. (Cronbach's $\alpha = .84$)

5. *Skilled Communicator*: Three-item measure of how effective at conveying lecture material the participant perceived the agent to be. Item reliability was questionable (Cronbach's $\alpha = .62$)

6. *Engagement*: Six-item measure of how engaged the participant felt during the lecture. Includes personal ratings of focus, attentiveness, and satisfaction. (Cronbach's $\alpha = .89$)

Manipulation Checks

To check that our *gaze type* manipulations were being noticed between the visible agent conditions, we asked the participants to rate from 0% to 100% how much they felt the agent was paying attention to them and to the map. We expected that participants would feel more attended to in the *affiliative* and *both* conditions than in the *referential* condition. Conversely, we expected participants to feel like the agent was attending to the map more in the *referential* and *both* gaze conditions than in the *affiliative* condition.

Results

Analysis of our data was conducted using a repeated measures analysis of variance (ANOVA). We used Tukey-Kramer HSD to control the experiment-wise error rate in all post-hoc tests. Our analysis started with the manipulation check. We found that participants felt more attended to in the *affiliative* gaze condition versus the *referential* condition, $F(1, 69) = 12.53, p < .001$. They also felt more attended to in the *both* condition versus *referential*, $F(1, 69) = 4.37, p = .040$. Finally, participants felt that the agent attended to the map more in the *referential* condition versus *affiliative*, $F(1, 69) = 7.75, p = .007$. This difference was not found to be significant for the *both* condition versus *affiliative*, $F(1, 69) = 0.37, p = .55$, however participants rated the agent as attending more to the map in the *referential* condition over the *both* condition, $F(1, 69) = 4.75, p = .033$.

Next we analyzed the objective results in the form of recall quiz scores (Figure 9). In terms of overall score, the *audio* condition resulted in significantly lower recall than the other three visible agent conditions, including *affiliative* gaze, $F(1, 69) = 19.38, p < .001$, *referential* gaze, $F(1, 69) = 37.78, p < .001$, and *both*, $F = 31.91, p < .001$. When considering only the seven (out of ten total) questions that dealt with purely spatial map information and building associations between verbal lecture content and locations on the map, we found that the *referential* gaze condition resulted in significantly better recall performance than the *affiliative* gaze condition, $F(1, 69) = 5.62, p = .021$. The increase in recall performance from the *both* condition over *affiliative* does not quite reach significance, $F(1, 69) = 2.58, p = .11$. *Referential* and *both* were not significantly different, $F(1, 69) = 0.59, p = .45$.

Finally, we analyzed the subjective measures. Here we observed that on the *likeability* scale, the *referential* condition rated lower than both the *affiliative* condition, $F(1, 69) = 58.86, p < .001$, and *both* condition, $F(1, 69) = 52.65, p < .001$. On the *rapport* scale, the *referential* condition also rated lower than both *affiliative*, $F(1, 69) = 13.25, p < .001$, and *both*, $F(1, 69) = 7.95, p = .006$. The *trust* scale had similar results, with the *referential* condition rated lower than both *affiliative*, $F(1, 69) = 8.63, p = .005$, and *both*, $F(1, 69) = 5.55, p = .021$. The *intelligence* scale again shows the *referential* condition getting rated lower than *affiliative*, $F(1, 69) = 11.38, p = .001$, and *both*, $F(1, 69) = 10.99, p = .002$. The *skilled communicator* scale yielded no significant results be-

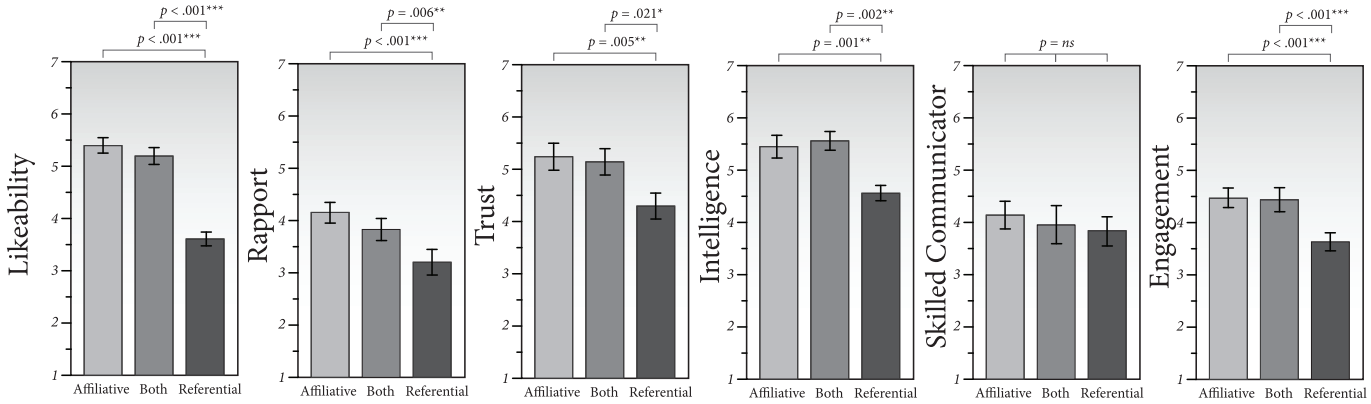

Figure 8. Results for subjective evaluations (likeability, rapport, trust, intelligence, skilled communicator, and engagement) based on gaze condition.

tween conditions, but the $engagement$ scale showed the *referential* condition getting rated significantly lower than the *affiliative* condition, $F(1, 69) = 14.53, p < .001$, and *both* condition, $F(1, 69) = 17.16, p < .001$. These results are summarized in Figure 8.

DISCUSSION

The purpose of the experiment was to demonstrate how subtle changes in gaze behavior can lead to significant high-level effects, with the goal of improving learning and positive feelings of affiliation. To do this, we manipulated the *head alignment* parameter in our model implemented on various virtual lecturing agents. We showed that by manipulating just this parameter for a virtual agent shifting its gaze from the participant to an object in the scene (i.e., the map of China) and back again, the agent could achieve very different subjective and objective effects, including the participant's feelings toward the agent and information recall respectively.

Confirming our first hypothesis, the mere presence of an agent resulted in better recall performance than audio alone, no matter which gaze condition was being used. For five out of six subjective scales, the *affiliative* and *both* gaze conditions achieved better ratings than the *referential* condition. This leads us to accept our second hypothesis, showing that human listeners prefer when the virtual agent speaker aligns its head fully with the participant while speaking, rather than looking out of the corners of its eyes with its head aligned towards something else. The *skilled communicator* scale did not exhibit this effect, which could mean that participants thought the agent was doing a similarly decent job of communicating the lecture content no matter which gaze condition was used. We also have strong support for our third hypothesis, where we showed that *referential* gaze results in better participant recall than *affiliative* gaze. By keeping its head aligned with the map as much as possible, the agent compelled the participant to concentrate more on the map and learn the spatial locations better, while building associations between verbal lecture content and those same locations. As a reminder, the same amount of gaze was used by the agent in both the *affiliative* and *referential* gaze conditions. Hence in both conditions the participant was able to benefit from gaze as an arousal stimulus to learn the lecture material better. The difference lies in the way each gaze shift was performed, proving that not all ways of gazing are equal and subtle changes can create significant outcomes. Overall, this study showed that our model was capable of achieving targeted outcomes.

Design Implications

Embodied agents' potential in computer interfaces is increasing with the ever-growing popularity of interactive games and computer-based tools for learning, motivation, and rehabilitation. However, agent designers need controllable models of interactive behavior and evidence that these models effectively improve outcomes such as learning and rapport. Our work provides one such example in the context of gaze.

We have shown that it is possible to implement in virtual agents the very subtle gaze cues that humans use to great effect in social situations, and that manipulating these cues can achieve significant objective and subjective high-level effects in a human interacting with the agent. Designers of virtual agents can use these subtle gaze behaviors, such as head alignment, to reach different desired outcomes. If the agent designer wants human interlocutors to pay more attention to specific objects in the environment, possibly to learn more about them, the agent could be programmed to use high head alignment when gazing to those objects. Similarly, if the agent designer wants the agent to build a stronger relationship with the human interlocutor, increasing feelings of, e.g., rapport and trust, the agent should be programmed to use high head alignment when gazing towards the human. Our model of gaze behavior offers a simple and effective means to control the low-level gaze parameters found in physiological research. Virtual agent designers can use and build off of this model to create rich, compelling gaze behaviors that accomplish the high-level effects they wish to achieve.

Limitations and Future Work

Our gaze mechanisms are currently focused on the contributions of the head and eyes alone. Future extensions should consider the movement of other body parts, such as employing the neck in performing head movements. While the work presented here explored our gaze shift model on very humanlike characters, we intend to explore its application to different character designs, e.g., stylized cartoons, and embodiments, e.g., storytelling robots [39]. We also intend to run future studies exercising different parameters of the model (e.g., head latency) and measuring the effects of subtle gaze cues in different interaction modalities, such as VR and mobile devices.

CONCLUSIONS

Gaze is a complex and powerful cue. Through subtle changes in gaze, people can achieve a wide range of social and communicative goals, affecting their partner of interaction in a variety of different ways. Gaze cues, as with all embodied communication cues, hold a strong fundamental connection with key social, cognitive, and task outcomes. This connection reveals an opportunity for designing embodied dialog with virtual agents. Designing gaze behaviors for virtual agents that can achieve specifically targeted high-level outcomes has long been a difficult problem. By creating a mechanism for synthesizing gaze shifts in a natural, yet parameterized fashion, we have provided a building block for creating high-level social and communicative behaviors. In this paper we presented a model of gaze shifts validated to achieve humanlike gaze for virtual agents, and a study which shows that it can achieve interesting subjective and objective effects through manipulation of low-level gaze parameters. These subtle gaze cues are effective in the context of our gaze model due to the grounded physiological approach we took in designing it. By manipulating and combining them in different ways, we believe that virtual agents will soon have access to a rich new source of possible gaze behaviors, resulting in human-agent interactions that are more effective and rewarding.

ACKNOWLEDGMENTS

This research was supported by National Science Foundation award 1017952. We would like to thank Allison Terrell, Danielle Albers, and Erin Spannan for their help in creating the videos used in the experiments described in this paper.

REFERENCES

1. Argyle, M., and Cook, M. *Gaze and mutual gaze*. Cambridge University Press Cambridge, 1976.
2. Bailenson, J., Yee, N., Merget, D., and Schroeder, R. The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence: Teleoperators and Virtual Environments 15*, 4 (2006), 359–372.
3. Bayliss, A., Paul, M., Cannon, P., and Tipper, S. Gaze cuing and affective judgments of objects: I like what you look at. *Psychonomic bulletin & review 13*, 6 (2006), 1061–1066.
4. Beebe, S. Effects of eye contact, posture and vocal inflection upon credibility and comprehension.
5. Burgoon, J., Coker, D., and Coker, R. Communicative effects of gaze behavior. *Human Communication Research 12*, 4 (1986), 495–524.
6. Cafaro, A., Gaito, R., and Vilhjálmsson, H. Animating idle gaze in public places. In *Intelligent Virtual Agents*, Springer (2009), 250–256.
7. Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsson, H., and Yan, H. Embodiment in conversational interfaces: Rea. In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, ACM (1999), 520–527.
8. Cassell, J., Torres, O., and Prevost, S. Turn taking vs. discourse structure: How best to model multimodal conversation. *Machine conversations* (1999), 143–154.
9. Deng, Z., Lewis, J., and Neumann, U. Automated eye motion using texture synthesis. *IEEE Computer Graphics and Applications* (2005), 24–30.
10. Fox, J., and Bailenson, J. Virtual virgins and vamps: The effects of exposure to female characters sexualized appearance and gaze in an immersive virtual environment. *Sex roles 61*, 3 (2009), 147–157.
11. Freedman, E., and Sparks, D. Activity of cells in the deeper layers of the superior colliculus of the rhesus monkey: evidence for a gaze displacement command. *Journal of neurophysiology 78*, 3 (1997), 1669.
12. Frischen, A., Bayliss, A., and Tipper, S. Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological bulletin 133*, 4 (2007), 694.
13. Fuller, J. Head movement propensity. *Experimental Brain Research 92*, 1 (1992), 152–164.
14. Fullwood, C., and Doherty-Sneddon, G. Effect of gazing at the camera during a video link on recall. *Applied Ergonomics 37*, 2 (2006), 167–175.
15. Garau, M., Slater, M., Bee, S., and Sasse, M. The impact of eye gaze on communication using humanoid avatars. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (2001), 309–316.
16. Goldberg, G., Kiesler, C., and Collins, B. Visual behavior and face-to-face distance during interaction. *Sociometry* (1969), 43–53.
17. Goldring, J., Dorris, M., Corneil, B., Ballantyne, P., and Munoz, D. Combined eye-head gaze shifts to visual and auditory targets in humans. *Experimental brain research 111*, 1 (1996), 68–78.
18. Goossens, H., and Opstal, A. Human eye-head coordination in two dimensions under different sensorimotor conditions. *Experimental Brain Research 114*, 3 (1997), 542–560.
19. Griffin, Z. Gaze durations during speech reflect word selection and phonological encoding. *Cognition 82*, 1 (2001), B1–B14.
20. Guitton, D., and Volle, M. Gaze control in humans: eye-head coordination during orienting movements to targets within and beyond the oculomotor range. *Journal of neurophysiology 58*, 3 (1987), 427.
21. Harris, M., and Rosenthal, R. No more teachers dirty looks: Effects of teacher nonverbal behavior on student outcomes. *Applications of nonverbal communication* (2005), 157–192.
22. Heylen, D., Van Es, I., Van Dijk, E., NI-JHOLT, A., van Kuppevelt, J., Dybkjaer, L., and Bernsen, N. Experimenting with the gaze of a conversational agent, 2005.
23. Hietanen, J. Does your gaze direction and head orientation shift my visual attention? *Neuroreport 10*, 16 (1999), 3443.
24. Ipeirotis, P. Demographics of Mechanical Turk. Tech. Rep. CeDER-10-01, 2010. Accessed on 10-Mar-2010 at http://hdl.handle.net/2451/29585.
25. Itti, L., Dhavale, N., and Pighin, F. Photorealistic attention-based gaze animation. In *2006 IEEE International Conference on Multimedia and Expo*, IEEE (2006), 521–524.
26. Kelley, D., and Gorham, J. Effects of immediacy on recall of information. *Communication Education* (1988).
27. Kim, K., Brent Gillespie, R., and Martin, B. Head movement control in visually guided tasks: Postural goal and optimality. *Computers in Biology and Medicine 37*, 7 (2007), 1009–1019.
28. Kittur, A., Chi, E., and Suh, B. Crowdsourcing user studies with Mechanical Turk. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, ACM (2008), 453–456.
29. Lance, B., and Marsella, S. The expressive gaze model: Using gaze to express emotion. *Computer Graphics and Applications, IEEE 30*, 4 (2010), 62–73.
30. Langton, S., and Bruce, V. Reflexive visual orienting in response to the social attention of others. *Visual Cognition* (1999).
31. Lee, J., Marsella, S., Traum, D., Gratch, J., and Lance, B. The rickel gaze model: A window on the mind of a virtual human. In *Intelligent Virtual Agents*, Springer (2007), 296–303.
32. Lee, S., Badler, J., and Badler, N. Eyes alive. In *ACM Transactions on Graphics (TOG)*, vol. 21, ACM (2002), 637–644.
33. Lester, J., Towns, S., Callaway, C., Voerman, J., and FitzGerald, P. Deictic and emotive communication in animated pedagogical agents. *Embodied conversational agents* (2000), 123–154.
34. Liu, C., Kay, D., and Chai, J. Awareness of partners eye gaze in situated referential grounding: An empirical study. *2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction* (2011).
35. Ma, X., Le, B., and Deng, Z. Perceptual analysis of talking avatar head movements: A quantitative perspective. *CHI'11* (2011).
36. Mason, M., Tatkow, E., and Macrae, C. The look of love: Gaze shifts and person perception. *Psychological Science* (2005), 236–239.
37. Mehrabian, A. Immediacy: An indicator of attitudes in linguistic communication. *Journal of Personality 34*, 1 (1966), 26–34.
38. Meyer, A., Sleiderink, A., and Levelt, W. Viewing and naming objects: Eye movements during noun phrase production. *Cognition 66*, 2 (1998), B25–B33.
39. Mutlu, B., Forlizzi, J., and Hodgins, J. A storytelling robot: Modeling and evaluation of human-like gaze behavior. In *Humanoid Robots, 2006 6th IEEE-RAS International Conference on*, IEEE (2006), 518–523.
40. Nijholt, A., Heylen, D., and Vertegaal, R. Inhabited interfaces: Attentive conversational agents that help. In *Proceedings 3rd international Conference on Disability, Virtual Reality and Associated Technologies-CDVRAT2000, Alghero, Sardinia* (2000).
41. Otteson, J., and Otteson, C. Effect of teacher's gaze on children's story recall. *Perceptual and Motor Skills* (1980).
42. Parke, F., and Waters, K. *Computer facial animation*. AK Peters Ltd, 2008.
43. Pelachaud, C., and Bilvi, M. Modelling gaze behavior for conversational agents. In *Intelligent Virtual Agents*, Springer (2003), 93–100.
44. Pelz, J., Hayhoe, M., and Loeber, R. The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research 139*, 3 (2001), 266–277.
45. Peters, C. Animating gaze shifts for virtual characters based on head movement propensity. In *2010 Second International Conference on Games and Virtual Worlds for Serious Applications*, IEEE (2010), 11–18.
46. Rickel, J., and Johnson, W. Task-oriented collaboration with embodied agents in virtual worlds. *Embodied conversational agents* (2000), 95–122.
47. Sherwood, J. Facilitative effects of gaze upon learning. *Perceptual and Motor Skills* (1987).
48. Steptoe, W., and Steed, A. High-fidelity avatar eye-representation. In *Virtual Reality Conference, 2008. VR'08. IEEE*, IEEE (2008), 111–114.
49. Tartaro, A., and Cassell, J. Authorable virtual peers for autism spectrum disorders. In *Proceedings of the Combined workshop on Language-Enabled Educational Technology and Development and Evaluation for Robust Spoken Dialogue Systems at the 17th European Conference on Artificial Intellegence*, Citeseer (2006).
50. Vertegaal, R. The gaze groupware system: mediating joint attention in multiparty communication and collaboration. In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, ACM (1999), 294–301.
51. Zangemeister, W., and Stark, L. Types of gaze movement: variable interactions of eye and head movements. *Experimental Neurology 77*, 3 (1982), 563–577.