Gaze and Attention Management for Embodied Conversational Agents

TOMISLAV PEJSA, SEAN ANDRIST, MICHAEL GLEICHER, and BILGE MUTLU,
University of Wisconsin-Madison

3

To facilitate natural interactions between humans and embodied conversational agents (ECAs), we need to endow the latter with the same nonverbal cues that humans use to communicate. Gaze cues in particular are integral in mechanisms for communication and management of attention in social interactions, which can trigger important social and cognitive processes, such as establishment of affiliation between people or learning new information. The fundamental building blocks of gaze behaviors are *gaze shifts*: coordinated movements of the eyes, head, and body toward objects and information in the environment. In this article, we present a novel computational model for gaze shift synthesis for ECAs that supports parametric control over coordinated eye, head, and upper body movements. We employed the model in three studies with human participants. In the first study, we validated the model by showing that participants are able to interpret the agent's gaze direction accurately. In the second and third studies, we showed that by adjusting the participation of the head and upper body in gaze shifts, we can control the strength of the attention signals conveyed, thereby strengthening or weakening their social and cognitive effects. The second study shows that manipulation of eye–head coordination in gaze enables an agent to convey more information or establish stronger affiliation with participants in a teaching task, while the third study demonstrates how manipulation of upper body coordination enables the agent to communicate increased interest in objects in the environment.

Categories and Subject Descriptors: H.1.2 [**User/Machine Systems**]: Human Factors; H.5.2 [**User Interfaces**]: User-centered design; I.3.7 [**Three-Dimensional Graphics and Realism**]: Animation

General Terms: Design, Experimentation, Human Factors, Animation

Additional Key Words and Phrases: Embodied conversational agents, gaze model, attention, body orientation, learning, affiliation

ACM Reference Format:
Tomislav Pejsa, Sean Andrist, Michael Gleicher, and Bilge Mutlu. 2015. Gaze and attention management for embodied conversational agents. ACM Trans. Interact. Intell. Syst. 5, 1, Article 3 (March 2015), 34 pages.
DOI: http://dx.doi.org/10.1145/2724731

1. INTRODUCTION

Embodied conversational agents have great potential as user interfaces across a range of application domains including education, health care, service industries, and entertainment. This potential stems from their ability to facilitate natural interaction with humans through the use of communicative cues, such as speech, facial expressions,

---

The reviewing of this article was managed by associate editor Elisabeth André.
Parts of this work were published in Proceedings of CHI 2012 [Andrist et al. 2012a] and Proceedings of the 4th ICMI Workshop on Eye Gaze in Intelligent Human Machine Interaction [Andrist et al. 2012b]. This work was supported by National Science Foundation award 1017952.
Authors' addresses: T. Pejsa, S. Andrist, M. Gleicher, and B. Mutlu, University of Wisconsin-Madison, Department of Computer Sciences, 1210 W. Dayton St., Madison, WI 53706-1613; emails: {tpejsa, sandrist, gleicher, bilge}@cs.wisc.edu.

DOI: http://dx.doi.org/10.1145/2724731

gaze, and gestures. By employing such cues in interaction, agents can trigger in people social and cognitive processes that lead to improved task outcomes. It has been shown that an agent employing a range of communicative cues can motivate people [Mumm and Mutlu 2011], help them learn better [Lusk and Atkinson 2007], and teach them social skills [Tartaro 2007]. For this reason, we aim to construct computational models of human communicative behaviors and study how these models might be employed to build more effective ECAs.

In this work, we focus on modeling *gaze shifts*: coordinated movements of the eyes, head, and body toward objects and information in the environment [Zangemeister and Stark 1982; McCluskey and Cullen 2007]. Gaze shifts serve as fundamental units of gaze behavior, which itself plays a key role in human communication. In particular, gaze is used to convey and manage attention in social interactions between people and enable beneficial social and cognitive processes. The goal of the current work is to enable designers to compose agents' gaze shifts into larger gaze behaviors that effectively manage people's attention and thus trigger these social and cognitive processes.

To illustrate the types of effects we seek to enable in human interactions with ECAs, we consider a scenario during which two humans engage in a conversation. As they talk, they are likely to gaze into each other's eyes, a state referred to as *mutual gaze* or eye contact. Mutual gaze creates a sense of being attended to and triggers positive social and cognitive effects: the two individuals engaging in eye contact will see each other as more likeable and truthful [Goldberg et al. 1969; Argyle and Cook 1976], and they will also be more likely to recall information discussed during the conversation [Otteson and Otteson 1980; Sherwood 1987]. As their interaction progresses, one person might make verbal references to an object in their shared environment, for example, a map. While doing so, the person will gaze at the map, signaling to the other person that his or her attention has shifted to the map. The other person's own gaze will then shift to the map in a seemingly automatic fashion [Frischen et al. 2007]. These two individuals are said to be engaging in *joint attention* [D'Entremont et al. 2007] via a shared focus on the map, which facilitates the learning of information by building associations between verbal references and the environment [Woodward 2005]—in our example, the second person may recall locations from the map as a result of joint attention.

The work presented in this article consists of two components. The first component is a novel computational model for synthesis of an agent's gaze shifts, which we describe in Section 3. The model is characterized by two key properties. First, it is a procedural animation model informed by neurophysiological measurements of gaze that enable it to synthesize natural-looking gaze shifts without the use of motion capture data. Second, the model can generate not only eye motion but also coordinated motion of the eyes, head, and upper body. Much of the prior research on gaze modeling for ECAs has focused only on eye gaze; however, research in neurophysiology and the social sciences has shown that a person tends to also reorient his or her head and body when shifting attention toward an object or person of interest and that an observer infers the direction of this attention not only from the eyes but also by aggregating information about the eye, head, and body orientation of the other individual [Perrett and Emery 1994; Hietanen 1999, 2002; Pomianowska et al. 2011]. For example, if a person gazes at an observer "out of the corner of the eye" while his or her head remains turned in another direction, the observer will respond to this attention cue differently than if the person completely turned his or her eyes and head in the observer's direction [Hietanen 1999]. Our model supports a broad range of attention cues by incorporating coordinated movements of the eyes, head, and upper body. These movements are parametrically controllable—the model exposes a set of *alignment parameters*, which the designer can use to define how much the agent's head and trunk will participate in a gaze shift. Figure 1 illustrates the usage of these parameters to produce significantly different

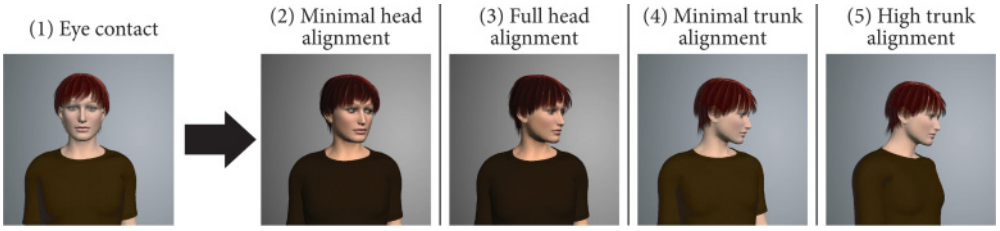

Fig. 1. Examples of gaze shifts synthesized using our model: (1) Initially, the agent maintains eye contact with the observer. (2) Gaze shift to the side with low value of the head alignment parameter. (3) Gaze shift in the same direction, but with high head alignment value. (4) Gaze shift in the same direction with a small amount of trunk motion. (5) As before, but with a high amount of trunk motion.

gaze shifts. We validated the communicative accuracy and naturalness of the model in a study with human participants, which we label Study 1 and present in Section 4 of the article.

The second component of this work is a pair of studies with human participants (labeled Study 2 and Study 3, respectively) in which we examined how control over eye, head, and upper body coordination in gaze shifts enables agents to communicate more effectively. In these studies, the agent used gaze shifts generated by our gaze model to manage the participants' attention in order to better convey certain information or influence the participants' judgments of the agent. Study 2 is an educational task in which we manipulated the eye–head coordination of a virtual teaching agent to produce two distinct patterns of gaze behavior. One of these patterns emphasized information in the environment, thus facilitating establishment of reference and joint attention, which led to improved information recall, while the second pattern emphasized eye contact, which facilitated feelings of affiliation and led to more positive perceptions of the agent. The setting for Study 3 was a virtual art gallery, and we manipulated the amount of upper body movement in the agent's gaze. Gaze shifts with more upper body reorientation would signal a major shift in attention to an art exhibit in the environment and would lead to a stronger perception of interest in that exhibit. The common feature of both studies is the use of alignment parameters to control the amount of head and upper body movement in gaze. Increased movement of the head and upper body signaled a greater shift in attention and augmented the social and cognitive effects produced by the gaze cue.

In summary, our work makes the following contributions:

(1) A procedural, empirically validated model for gaze shift synthesis that enables parametric control over the coordination of eye, head, and upper body movements
(2) A demonstration, in a study with human participants (Study 2), of how control over the agent's eye–head coordination enables the agent to convey information more effectively or establish better affiliation with people interacting with the agent
(3) A demonstration, in a study with human participants (Study 3), of how control of the agent's upper body movements enables the agent to better communicate interest in nearby objects to people interacting with the agent

We emphasize that the latter two contributions are distinct from the model that makes up the first contribution. The studies in question were designed not merely to evaluate the current model but to explore the design space of ECA's gaze. Their value is in showing how more effective attention management mechanisms can be constructed from gaze shifts with varying coordination properties. In these studies, we utilized our own gaze model to synthesize such shifts, but we expect their results to generalize to other models that might offer a comparable degree of parametric control.

2. RELATED WORK

Gaze has been a subject of research in a number of academic disciplines. Social sciences have focused on studying the characteristics and functions of gaze in human communication, with the tendency to provide qualitative descriptions of gaze behaviors. In neurophysiology, there has been a focus on physical aspects of gaze behaviors, including quantitative data about kinematic properties of gaze shifts derived from highly controlled laboratory experiments. These two bodies of work are complementary and provide sufficient information to construct and evaluate computational models of gaze. The gaze model presented in this work is derived from quantitative measurements of gaze in neurophysiology, while the hypotheses about social and cognitive effects of gaze tested in our studies are informed by research in the social sciences.

In this section, we review the social sciences research that examines the role of gaze in human communication as well as neurophysiological measurements of gaze. In addition, we discuss prior work on gaze modeling for embodied conversational agents, including an overview of existing computational models of gaze and how such models have been employed to build more capable agents.

2.1. Gaze in Human Communication

The ability to follow another person's gaze and infer the direction of his or her attention emerges in the first year of life [Scaife and Bruner 1975]. Following another person's gaze enables establishment of reference [D'Entremont et al. 2007], inference of intent [Mumme et al. 2007], and engagement in joint attention with the person [D'Entremont et al. 2007]. Joint attention facilitates learning and comprehension by associating speech utterances with objects and information in the environment. This ability also develops at an early age and is crucial for the acquisition of language [Woodward 2005].

People are particularly sensitive to the gaze of others [Argyle and Cook 1976]. This is interpreted as being attended to and, depending on the context of interaction, can lead to either discomfort from feeling observed or genuine social interaction [Argyle and Cook 1976]. A person who makes increased eye contact is associated with greater perceived dynamism, likeability, and believability [Beebe 1976]. These people are also seen as being more truthful or credible [Argyle and Cook 1976]. One illustrative study specifically showed that people who spend more time gazing at an interviewer receive higher socioemotional evaluations [Goldberg et al. 1969]. On the other hand, gaze aversion produces consistently negative effects in impressions of attraction, credibility, and relational communication [Burgoon et al. 1986].

Gaze has been found to be a significant component of *immediacy*—defined as the degree of perceived physical or psychological closeness between people [Mehrabian 1966]—especially in the context of improving educational outcomes [Harris and Rosenthal 2005]. Students from primary school through college have been shown to learn better when they are gazed at by the lecturer [Otteson and Otteson 1980; Sherwood 1987]. 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 [Fullwood and Doherty-Sneddon 2006]. The positive effect of gaze on recall is usually attributed to its role as an arousal stimulus, which increases attentional focus and therefore enhances memory [Kelley and Gorham 1988].

Studies have shown that humans infer other people's direction of attention by integrating information about their eye gaze direction, head orientation [Hietanen 1999], and body orientation [Langton et al. 2000; Hietanen 2002; Pomianowska et al. 2011]. Changes in body orientation occur relatively rarely during interaction compared to gaze shifts involving eye and head movements and indicate larger shifts in attention

and mutual involvement [Kendon 1973; Schegloff 1998]. People tend to reorient their bodies to avoid head rotations larger than 90° when gazing at someone [Kendon 1973]. Changes in body orientation also occur when a new party joins a conversation in order to reconfigure the conversational formation [Kendon 2010].

These findings are relevant to our own work in two ways. First, they underscore the need for a gaze model for agents that incorporates coordinated eye, head, and upper body movements. Second, they inform our hypotheses regarding how agents' gaze behavior might lead to better interactions with humans.

2.2. Gaze Modeling in Neurophysiology

Research in neurophysiology has studied how humans and other primates carry out gaze shifts in a tightly connected dynamic process by coordinating eye, head, and body movements [Zangemeister and Stark 1982; André-Deshays et al. 1988; Barnes 1979; Freedman and Sparks 2000; Uemura et al. 1980; McCluskey and Cullen 2007]. Researchers measured these movements in highly controlled experiments, obtaining numeric data about kinematic properties such as movement range [Guitton and Volle 1987] and eye and head velocities [Guitton and Volle 1987; Freedman and Sparks 2000; Barnes 1979; Uemura et al. 1980]. The procedural gaze model presented in this article is based on their results.

Kinematics of coordinated eye and head movements are reported by Guitton and Volle [1987], Freedman and Sparks [2000], Barnes [1979], Uemura et al. [1980], and others. Eye and head movements in gaze are tightly coupled and significantly affect each other. There exists a linear relationship between head velocity and eye movement amplitude in gaze shifts [Barnes 1979; Uemura et al. 1980]. Furthermore, head movement amplitude decreases as eye movements start at increasingly contralateral positions (i.e., oriented away from the target in relation to head direction). Shifts that start at such positions require that the eyes contribute more to the shift [McCluskey and Cullen 2007]. The degree to which individuals use their heads in performing a gaze shift is highly idiosyncratic. Neurophysiological research literature describes some people as "head-movers," that is, individuals who move their head fully to align with the gaze target every time, and some as "non-head-movers" [Fuller 1992]. From a biomechanical standpoint, humans should universally be "non-head-movers," as fully moving the head—which is almost a hundred times heavier than the eyes—is not an economical solution [Kim et al. 2007]; however, the human tendency to move the head more than necessary during gaze shifts can be attributed to the role of head orientation in signaling attention [Hietanen 1999].

The reorientation of gaze direction is accomplished not only through movements of the eyes and the head but also through movements of the body. Unfortunately, studies that have attempted to measure the kinematics of these movements are few. McCluskey and Cullen [2007] measured upper body movements in primate gaze, finding that body contribution to gaze shifts increased with gaze shift amplitude and that the upper body followed eye and head movements with substantial latency. Based on these results, they concluded that body movements in gaze shifts are part of a series of coordinated motor events triggered when primates reorient their gaze. Hollands et al. [2004] obtained similar findings for coordination of the eyes, head, and feet during whole-body reorientation.

2.3. Gaze for Embodied Conversational Agents

Previous research on gaze for ECAs has focused on building computational models of human gaze behavior and studying how such models can improve human interactions with agents. In our review of this work, we make a distinction between models for synthesis of individual gaze shifts and high-level models that compose gaze shifts

into gaze behaviors. The former models synthesize eye and head motion toward a target specified manually by a designer or automatically by a high-level model of gaze behavior. The latter models do not simply synthesize the motion of the gaze shift; they also decide when gaze shifts should occur and toward which targets. We follow up the survey of gaze models with a review of work that has studied the effects of agent gaze in the context of interactions with humans, showing how appropriate gaze behaviors can lead to more positive perceptions of the agent and smoother, more productive interactions.

2.3.1. Computational Gaze Models. Models for gaze shift synthesis can be classified as data-driven models, which use motion capture data to synthesize novel gaze behaviors [Heck 2007; Lance and Marsella 2010], and procedural models, which synthesize gaze movements by employing kinematic laws [Andrist et al. 2012b; Peters and Qureshi 2010; Thiebaux et al. 2009]. Data-driven models can produce more natural gaze motions, while procedural models typically offer a greater degree of parametric control. The Expressive Gaze Model (EGM) [Lance and Marsella 2010] is an example of a hybrid model, which uses procedurally generated, neurophysiologically based eye movements in combination with motion-captured head and torso movements. In previous work, we introduced an earlier version of the gaze model we present in this work, which is a procedural model that supports parametric control over eye–head coordination [Andrist et al. 2012b]. The model by Peters and Qureshi [2010] is another procedural model that supports parametric control over eye and head movements, via a head propensity parameter, and is functionally similar to our head alignment parameter.

Most gaze models incorporate only eye and head movements, or eye movements alone. Of those gaze models that incorporate upper body movement, EGM [Lance and Marsella 2010] uses motion capture data to animate the torso during gaze movements, and the Parametric Gaze Map by Heck [2007] synthesizes gaze shifts by interpolating motion-captured poses, while the model by Grillon and Thalmann [2009] uses an IK solver to adapt the upper body pose toward a gaze target. None of these models provides parametric control over upper body coordination. To our knowledge, the only other model to provide parametric control over both head and upper body alignment is the Smart Body Gaze Controller (SBGC) [Thiebaux et al. 2009]. The SBGC uses a cyclic coordinate descent (CCD) solver to manipulate the chain of joints involved in the gaze shift while exposing a rich set of parameters for controlling joint velocities and their contributions to the gaze shift. Our model uses a similar approach to move gaze joints toward the target; the key difference is that we use neurophysiologically based kinematic laws to drive the motion. Moreover, our model is designed to produce neurophysiologically plausible motion across the entire parametric range, while alignment parameters are only manipulated to achieve specific communicative intent.

All of the gaze models discussed so far are low-level models for synthesis of individual gaze shifts. In addition, a number of high-level gaze models that generate entire sequences of gaze shifts have been presented. Timings and spatial directions of these shifts are usually chosen based on statistical models of human gaze behavior. This category includes models of gaze for turn management [Mutlu et al. 2012; Cassell et al. 1999; Pelachaud and Bilvi 2003], face-to-face interactions [Lee et al. 2002; Khullar and Badler 2001; Gu and Badler 2006; Andrist et al. 2013], and crowds [Grillon and Thalmann 2009; Cafaro et al. 2009] and models that trigger gaze shifts based on the agent's emotional or cognitive state [Li et al. 2009; Lee et al. 2007].

2.3.2. Gaze and Communication with ECAs. An agent's gaze behavior is very important to providing rich interactions and can have great impact on task performance and perceptions of the agent. A virtual agent employing gaze at turn-taking boundaries is evaluated more positively, and interactions with the agent proceed more efficiently

[Heylen et al. 2002]. By initiating and breaking eye contact at appropriate times, a conversational agent can appear as more thoughtful, can achieve smoother turn taking, and can elicit more disclosure from human participants [Andrist et al. 2013]. A virtual agent can also employ gaze as an attention cue and assist people in localizing task-relevant objects in the environment [Bailly et al. 2010]. Positive effects of gaze have also been demonstrated on physical embodiments, such as robots; for example, increased gaze from a storytelling robot facilitates greater recall of the story [Mutlu et al. 2006], while a conversational robot can utilize specific gaze patterns to clarify conversational roles in a multiparty interaction [Mutlu et al. 2012]. However, poor gaze behavior can be worse than no gaze at all. The positive effects of using an agent—as opposed to only audio or text—can be completely lost if the gaze is poor or random [Garau et al. 2001]. Further, the 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 [Fox and Bailenson 2009].

Previous research on the gaze of ECAs has mostly focused on creating high-level models of gaze behaviors, such as those discussed earlier, rather than manipulating properties of individual gaze shifts. Formal studies with human participants typically examined gross manipulation of the gaze behavior to show how the mere presence of the model can lead to improved agent perception and task performance. Our work instead focuses on studying how manipulations of specific gaze shift parameters can produce novel patterns of gaze behavior that map to changes in task outcomes and subjective evaluations of the agent.

3. GAZE MODEL

In this section, we describe our model for gaze shift synthesis for ECAs. The model falls into the category of procedural, low-level models that synthesize individual gaze shifts toward targets in space. It is distinguished from prior models in that it offers intuitive parametric control over eye, head, and upper body coordination while synthesizing neurophysiologically plausible motion across the parametric range. The gaze shifts to synthesize can be specified manually by a designer, or an automated, high-level gaze model could utilize our model for synthesis of individual gaze shifts, taking advantage of its rich parameterization to produce a broader range of gaze behaviors.

The model accepts as input the properties of the gaze shift that should be synthesized, including *gaze target position* (where the agent will look) and head and trunk *alignments*, which specify how far each of these body parts will rotate relative to the target. Given these inputs, the model synthesizes a gaze shift that turns the agent's eyes—and potentially also head and upper body—toward the target. The animation of the gaze shift is generated using a set of kinematic laws, derived from measurements of primate gaze reported in neurophysiology research [Guitton and Volle 1987; McCluskey and Cullen 2007]. Articulation is achieved by incrementally rotating the joints of the eyes, neck, and lower spine toward the target; rotational increments are computed on every frame using the kinematic laws, based on the current target position relative to the agent. The approach enables the model to synthesize accurate gaze shifts even toward targets that are moving relative the agent.

Assuming the agent is a virtual character, our model only requires them to be rigged for skeletal animation and to have the appropriate set of joints. Given a rigged character with humanlike anatomic proportions, the gaze model can be used as is—although kinematic laws utilized in the model expose a number of designer-tunable constants, these do not need to be adjusted to synthesize neurophysiologically correct gaze shifts. Furthermore, in order to simplify parametric control and make the model generalize better to different characters, body joints are distributed into several groups that are controlled jointly. Neck joints (cervical vertebrae) are grouped together under the body

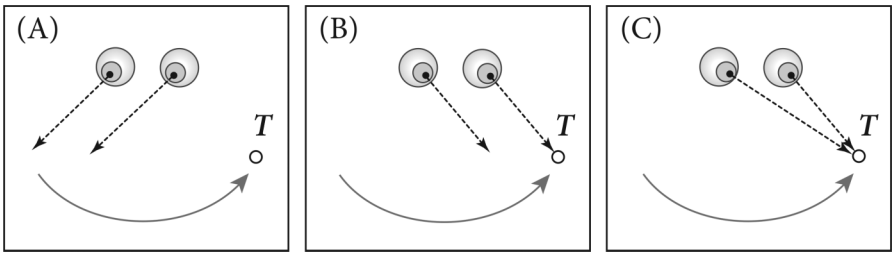

Fig. 2. Phases of an eye saccade. Dashed arrows indicate eye gaze directions, while the curved arrow indicates the direction of the rotational movement. Saccade proceeds as follows: (A) Eyes begin to rotate toward the target. (B) First eye has reached the target. (C) Both eyes have reached the target.

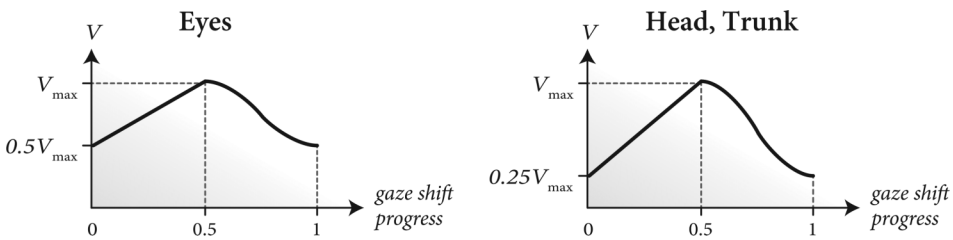

Fig. 3. Velocity profiles for the motion of different body parts. Left: Eye velocity. Right: Head and trunk velocities. Mathematical expressions for the profiles are given in Equations (9) and (10).

part *head*. As the head rotates over the course of a gaze shift, this rotation is distributed among the joints that constitute the head. This is based on how real humans move—head movement in humans is achieved through simultaneous rotation of all the neck joints. Similarly, trunk rotation is achieved through simultaneous rotation of the lower spine joints (lumbar vertebrae), and so these are all grouped under the body part *trunk*. In the article, when we refer to head rotation or trunk rotation, we actually refer to rotation that is distributed among the joints of the neck and lower spine.

3.1. Eye Movements

In this section, we describe eye movements that occur in gaze shifts as they are simulated in our model. Directed eye movements in gaze shifts are referred to as *saccades*. Saccades are realized as shortest-path rotations of the eyes in their sockets toward the gaze target T. Position of the target, T, is a designer-specified parameter of the model. Figure 2 illustrates a saccadic eye movement. At the start (A), the eyes begin to rotate toward the gaze target simultaneously. One of the eyes reaches the target first and locks onto it (B), while the other eye continues to rotate toward the target until it has aligned with it as well (C). The opposing eye movements that occur between (B) and (C) are referred to as *vergence*. We use the same kinematic model to produce both saccadic and vergence motions, although neurophysiology research has shown these are neurally distinct motions that activate different, albeit substantially overlapping areas of the brain [Alkan et al. 2011].

At the end of the gaze shift, both eyes are locked onto the gaze target. If the relative position of the target and the eyes changes due to head motion, the eyes automatically move in the opposite direction to compensate and maintain their lock on the target. This eye movement is known as the *vestibulo-ocular reflex* (VOR) and its function is to stabilize the image on the eye's retina during movement.

Saccadic motion of the eyes follows the piecewise polynomial velocity profile shown in Figure 3, left. Peak eye velocity $V_{\max,E}$ is computed at the start of the gaze shift

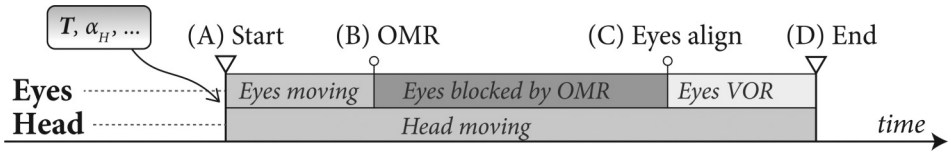

Fig. 4. Movement phases of the eyes and head in a gaze shift.

from the amplitude of the upcoming eye movement. This is in accordance with human gaze shifts, which take a similar amount of time to complete, irrespective of amplitude. Human saccades are extremely fast—peak eye velocity can be as high as $900^\circ/s$ for large gaze shifts [Bahill et al. 1975].

We compute the peak eye velocity using the following formula, derived from measurements of gaze kinematics reported by Guitton and Volle [1987]:

$$V_{\max,E} = \left(\frac{2}{75}A_{\min} + \frac{1}{6}\right)V_{0,E} \tag{1}$$

$$A_{\min} = min\left(\underset{j\in\text{Eyes}}{min}\,(D_j), \text{OMR}\right).$$

$V_{0,E}$ is an overridable eye-velocity parameter, which we keep at the default value of $150^\circ/s$, while $A_{\min}$ is the estimated amplitude of the eye movement. Eyes are mechanically restricted in how far they can rotate in their sockets. The maximum range of eye movement is referred to as *ocular motor range* (OMR), and it is between 45° and 55° in any direction. When the gaze target lies beyond the eyes' motor range, we set eye movement amplitude $A_{\min}$ to be equal to OMR value; otherwise, we set it to the lower of the angular distances D_j between initial eye gaze direction and target direction.

3.2. Eye–Head Coordination

When gazing at a target that lies beyond the eyes' OMR, some head movement must occur during the gaze shift to enable the eyes to reach the target. However, as discussed in Section 2.2, head movements do not occur solely for mechanical reasons; otherwise, the head would always rotate only the minimal amount needed to see the target. Gaze shifts often involve much greater head movements because head orientation is part of the human attention cueing mechanism [Hietanen 1999]. The amount of head movement in a gaze shift therefore has important implications for how the gaze shift is perceived. One of the key aspects of our model is capturing the coupling between eye and head movements during gaze shifts.

In this section, we describe how the coordinated eye and head movements are implemented in the model. At the start of a gaze shift, the designer specifies the values of two control parameters: gaze target position in world space and head alignment α_H. Head alignment specifies how much the head should participate in the gaze shift. At $\alpha_H = 0$, the head will rotate only the minimal amount needed to permit the eyes to see the target (so the agent will look at the target "out of a corner of the eye"). For $\alpha_H = 1$, the head will be fully aligned, causing the agent to completely face the target.

The gaze shift consists of several phases (Figure 4). At the start (A), both the eyes and head begin rotating toward the target. Since the eyes move much faster than the head, they quickly reach either the target or the limit of their OMR, set to be between 45° and 55° in any direction (B). There they remain blocked until the head catches up. Eventually the head rotates far enough to bring the eyes into alignment with the target (C). The VOR locks the eyes onto the target as the head continues to align. Depending

on the specified head alignment, the head will either stop moving immediately ($\alpha_H = 0$) or continue rotating until it has reached the required alignment with the target (D).

Throughout the entire gaze shift, head motion follows a piecewise polynomial velocity profile similar to the one for the eyes (Figure 3, right). The peak head velocity $V_{\max,H}$ is computed at the start of the gaze shift; it is proportional to the expected amplitude of the head movement [Guitton and Volle 1987]:

$$V_{\max,H} = \left(\frac{4}{3}\frac{D_H}{50} + \frac{2}{5}\right) V_{0,H}. \tag{2}$$

D_H is rotational amplitude of the head or distance it will rotate over the course of the gaze shift. Velocity parameter $V_{0,H}$ is designer specifiable, with the default value of $50^\circ/s$.

3.2.1. OMR Estimation. As mentioned, human eyes are mechanically limited in their movements by the OMR, which is estimated to be between 45° and 55°. Encoding these OMR values as static parameters into virtual humans is not sufficient, however, as the effective OMR depends on the initial position of the eyes at the start of the gaze shift and may fluctuate over the course of the gaze shift. These fluctuations are a product of a neural (as opposed to mechanical) limitation imposed on eye motion [Guitton and Volle 1987]. We define initial eye position (IEP) as rotational offset of the eyes from the central orientation (when the eyes are looking "straight ahead"). We define IEP to be nonzero only when the eyes are *contralateral* to the target (i.e., their initial orientation is on the opposite side of the center from the target). The relationship between IEP and neurally limited OMR is obtained from Guitton and Volle [1987]:

$$\mathrm{OMR} = \mathrm{OMR}_0 \left(\frac{1}{360}\mathrm{IEP} + 0.75\right), \tag{3}$$

where OMR_0 is the mechanical OMR value (between 45° and 55° by default).

3.2.2. Head Latency. In human gaze shifts, the head typically begins to move with some delay relative to the start of eye movements. This delay, which we refer to as *head latency*, typically ranges between 0 and 100*ms* and depends on task properties and individual characteristics—factors including gaze shift amplitude, predictability and saliency of the target, vigilance of the subject, and whether the gaze shift is forced or natural [Pelz et al. 2001; Zangemeister and Stark 1982]. In rare cases, head latency can be negative; that is, the head starts moving *before* the eyes. Studies that evaluated correlations between head latency and target modality found that eyes tend to lead the head when people orient toward visual targets, whereas the head tends to lead when orienting toward auditory targets [Goldring et al. 1996; Goossens and Opstal 1997].

Our gaze model exposes head latency as an additional eye–head coordination parameter τ_H. By setting a value $\tau_H > 0$, the designer can designate the head to start moving after the eyes, while for $\tau_H < 0$, the head will start moving before the eyes during the start of the gaze shift. In the experiments presented in the current work, this parameter is held constant at its default value of zero.

3.3. Upper Body Coordination

Gaze shifts often incorporate upper body movements in addition to eye and head movements. Similar to changes in head orientation, changes in upper body orientation do not occur simply because they are necessary mechanically, but because they demonstrate large shifts in attention and mutual involvement [Kendon 1973].

To support upper body movements, we make several extensions to the model of eye–head coordination described earlier. Development of these extensions was largely

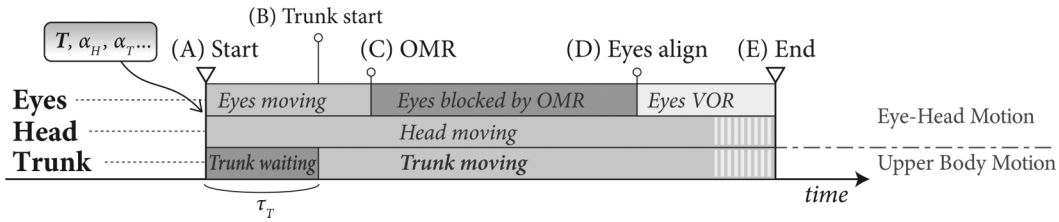

Fig. 5. Movement phases of the eyes, head, and trunk in a gaze shift. Dashed area indicates that the trunk may stop moving before or after the head.

guided by existing neurophysiological measurements of gaze and our own experimentation. We mostly referred to the study by McCluskey and Cullen [2007], who investigated eye, head, and body coordination during gaze shifts of rhesus monkeys with the expectation that the results would apply to other species of primates.

Thus, we introduce a new body part—the trunk, which has its own alignment parameter α_T. Trunk alignment specifies the participation of the trunk in the gaze shift. At $\alpha_T = 0$, the trunk will rotate only the minimal amount, while $\alpha_T = 1$ will make the trunk fully align with the target.

It has been shown that the amount of trunk movement increases with gaze shift amplitude [McCluskey and Cullen 2007]. Small gaze shifts do not engage the trunk at all. The minimal amount of trunk rotation as a function of angle in degrees is computed as follows:

$$D_{T,\min} = \begin{cases} 0.43\,exp(0.03D) + 0.19 & \text{iff } D \geq 40^\circ \\ 0.08D - 1.56 & \text{iff } 20^\circ \leq D < 40^\circ \\ 0 & \text{iff } D < 20^\circ, \end{cases} \tag{4}$$

where D is the overall gaze shift amplitude, estimated as the angle between current and target eye gaze direction. Note that for gaze shifts where $D \geq 20^\circ$, the trunk will rotate by the amount $D_{T,\min}$ even if $\alpha_T = 0$. For higher values of α_T, the trunk will rotate toward the target even more.

The sequence of phases in an upper body gaze shift is shown in Figure 5. At the start of the gaze shift (A), the trunk is stationary for a period of τ_T *ms*, after which it begins to move. The trunk continues to move until it reaches the correct orientation relative to the target, as specified by α_T. The trunk moves more slowly than the eyes and head, so it may continue to move even as the eyes and head reach and lock onto the target (D, E). However, for gaze shifts with low latency and little trunk movement, the trunk may reach its final orientation while the head is still moving.

Since prior studies of upper body motion do not provide an exact profile of trunk velocity, we apply the peaked profile used for head velocity to trunk motion as well (Figure 3, right). Peak trunk velocity $V_{\max,T}$ should be lower than peak head velocity, but it should still be proportional to rotational amplitude [McCluskey and Cullen 2007]. Guided by these principles and experimentation, we derived the following expression for the peak trunk velocity:

$$V_{\max,T} = \left(\frac{4}{3}\frac{D_j}{15} + 2\right) V_{0,T}, \tag{5}$$

where D_j is the rotational amplitude of the trunk, while $V_{0,T}$ is a designer-specifiable velocity value, with the default value of $15^\circ/s$.

3.3.1. Trunk Latency. McCluskey and Cullen [2007] found that the trunk does not begin to move at the same time as the eyes and head, but with some latency τ_T that depends

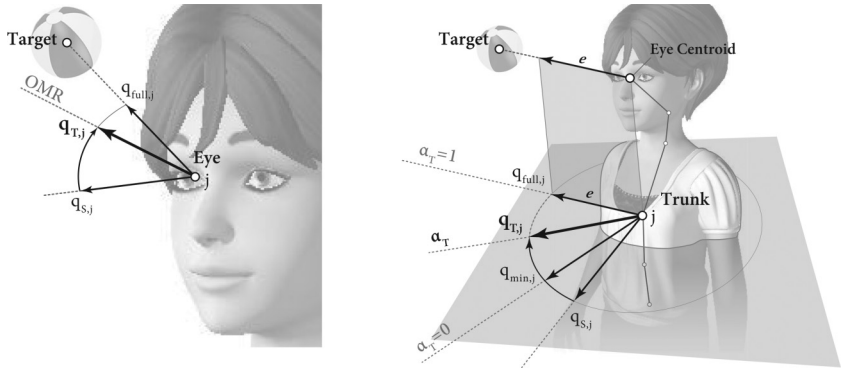

Fig. 6. Computing target orientation $q_{T,j}$ for a joint j. Left: Procedure for the eyes. $q_{\text{full},j}$ is the orientation that fully aligns the eye with the gaze target. However, since the target lies beyond the eye's OMR, $q_{\text{full},j}$ is clamped to the OMR limit, giving us the target orientation $q_{T,j}$. Right: Procedure for trunk joints. $q_{\text{min},j}$ is the minimal orientation the trunk must achieve, while $q_{\text{full},j}$ is the orientation that fully aligns the trunk with the gaze target. The latter is given by the vector e pointing from the eye centroid to the target. Target orientation $q_{T,j}$ is computed by interpolating between minimal and fully aligned orientations using the alignment parameter α_T.

on factors such as gaze shift amplitude and target predictability. In our experiments, we found that trunk latency value had great impact on naturalness of upper body motion, so we could not simply set it to a fixed value. Instead, based on the results from the McCluskey study, we compute trunk latency (in milliseconds) as follows:

$$\tau_T = 0.25D + 47.5, \tag{6}$$

where D is gaze shift amplitude. It follows from this equation that in large gaze shifts, the trunk is going to trail behind the head and eyes more.

3.4. Animating a Gaze Shift

The parameters described previously specify a complete gaze shift. Motion synthesis is accomplished through feed-forward control, which drives the joints of the eyes, head, and trunk toward the target at angular velocities computed from their velocity profiles. Each joint has a numerical index j: left and right eye have joint indices 0 and 1, respectively, while joints lower in the body (head, trunk) are assigned incrementally higher indices. Joint orientations are represented using quaternions—four-dimensional numbers that encode the angle and axis of a 3D orientation. We denote the orientation of the joint j at the start of the gaze shift $q_{S,j}$. We also define the *target orientation* $q_{T,j}$, which is the final orientation of the joint at the end of the gaze shift. As the gaze shift progresses, the joint rotates from the source orientation to the target orientation—its current orientation gets updated on every animation frame. We denote the current orientation of joint j at frame i as q_j^i.

3.4.1. Computing Target Joint Orientations. Before the gaze shift begins, we compute the target orientation for each joint $q_{T,j}$. For the eyes, the computation proceeds as follows (Figure 6, left). We compute the fully aligned eye orientation $q_{\text{full},j}$—the orientation at which the eye is staring directly at the target. If the target lies within the eye's OMR (i.e., if it is reachable by the eyes), then $q_{T,j} = q_{\text{full},j}$. If not, then $q_{T,j}$ is obtained by clamping $q_{\text{full},j}$ to the OMR limit (computed using Equation (3)).

The method of computing target orientations for trunk joints is illustrated in Figure 6 (right). We first compute the fully aligned trunk orientation $q_{\text{full},j}$, which is the orientation the trunk joint would have if it were completely facing the target. We choose $q_{\text{full},j}$

such that the facing direction of the joint j at that orientation is parallel to the vector e pointing from the eye centroid to the target. We also know the trunk must rotate by a minimal amount $D_{T,\min}$ (see Equation (4)), which we use to compute minimal trunk orientation $q_{\min,j}$. Target orientation lies between the minimal orientation $q_{\min,j}$ and fully aligned orientation $q_{\text{full},j}$, depending on the trunk alignment parameter α_T. We can compute it using spherical linear interpolation (*slerp*), an operation that interpolates a pair of quaternions along the shortest arc, using an interpolation parameter that ranges from 0 to 1. We use *slerp* to interpolate between $q_{\min,j}$ and $q_{\text{full},j}$ using trunk alignment α_T as the interpolation parameter:

$$q_{T,j} = slerp(q_{\min,j}, q_{\text{full},j}, \alpha_T). \tag{7}$$

$q_{\min,j}$ lies between the source orientation $q_{S,j}$ and the fully aligned orientation $q_{\text{full},j}$ and can also be computed using *slerp*:

$$q_{\min,j} = slerp\left(q_{S,j}, q_{\text{full},j}, \frac{D_{T,\min}}{\angle(q_{S,j}, q_{\text{full},j})}\right), \tag{8}$$

where the symbol $\angle$ denotes the angular distance between two quaternions.

The method for head joints is nearly identical to that for trunk joints (Figure 6, right), with two differences. First, we use the head alignment parameter α_H instead of trunk alignment α_T. Second, $q_{\min,j}$ is the minimal head orientation needed to bring the eyes into alignment with the gaze target; that is, it is computed from the highest rotational difference between each eye's $q_{\text{full},k}$ and OMR-clamped $q_{T,k}$ ($k \in$ Eyes).

3.4.2. Updating Joint Orientations. As the gaze shift executes, joint orientations are updated at the constant frame rate of 60*fps*. On every frame, the update algorithm iterates through all the joints in the hierarchy, starting at the lowermost trunk joint and ending with the eyes. If the joint is a trunk joint or head joint, the algorithm checks whether its latency time (τ_T or τ_H, respectively) has expired—if it has, the joint is ready to begin rotating. We compute the joint's current rotational velocity V_j^i, where i is the current frame index, using the appropriate velocity profile. Then we incrementally rotate the joint toward its target orientation $q_{T,j}^i$. Rotating a joint affects the target orientations of subsequent joints ($q_{T,j-1}^i$, $q_{T,j-2}^i$, etc.), which therefore need to be recomputed. The gaze shift ends when both eyes have aligned with the target and all head and trunk joints have reached their target orientations.

The velocity profiles used to compute V_j^i for each joint are shown in Figure 3. We derived these piecewise polynomial profiles as approximations of eye and head velocity profiles reported in the literature [Kim et al. 2007; Lee et al. 2002]. The velocity profile for the eyes has the following mathematical expression:

$$V_j^i = \begin{cases} (r_j^i + 0.5)V_{\max,E} & \text{iff } r_j^i < 0.5 \\ (8{r_j^i}^3 - 18{r_j^i}^2 + 12r_j^i - 1.5)V_{\max,E} & \text{iff } r_j^i \geq 0.5. \end{cases} \tag{9}$$

The velocity profile for the head and trunk (Figure 3, right) is defined as follows:

$$V_j^i = \begin{cases} (1.5r_j^i + 0.25)V_{\max,j} & \text{iff } r_j^i < 0.5 \\ (12{r_j^i}^3 - 27{r_j^i}^2 + 18r_j^i - 2.75)V_{\max,j} & \text{iff } r_j^i \geq 0.5. \end{cases} \tag{10}$$

r_j^i is the gaze shift progress parameter, which monotonically increases from 0 to 1 over the course of the gaze shift. Joint j is considered to have reached the target when $r_j^i = 1$. The parameter r_j^i is updated on every frame based on angular distance covered

by the joint. The update equation is as follows:

$$r_j^{i+1} = r_j^i + \frac{\Delta t \cdot V_j^i}{\angle(q_{S,j}, q_{T,j}^i)} c_j \tag{11}$$
$$c_j = \frac{2(N_j - j + o_j)}{N_j(N_j + 1)},$$

where Δt is time since the last update, N_j is the number of joints in the current body part, and o_j is the index of the topmost joint in the current body part. For example, the trunk might consist of three joints ($N_j = 3$) linked in a chain, with indices starting at $o_j = 4$ for the topmost joint and ending at 6 for the lowermost joint. As we update the orientation of each joint, we multiply its parameter increment by the *joint contribution* c_j to ensure that overall rotation of the body part is distributed among its joints. The joint contribution c_j will have a higher value for joints that have lower index j, meaning they are higher up in the body and must contribute more to the overall rotation of the body part. In our example with the three-joint trunk, contributions of each trunk joint will be $\frac{1}{2}$, $\frac{1}{3}$, and $\frac{1}{6}$, respectively. Note that for the eyes, which are not joint chains but single joints, we always have $N_j = 1$ and hence $c_j = 1$.

Having computed the rotation progress r_j^{i+1} for each joint, we can update its current orientation by interpolating between the starting orientation $q_{S,j}$ and the target orientation $q_{T,j}$:

$$q_j^{i+1} = slerp\left(q_{S,j}, q_{T,j}^i, r_j^{i+1}\right). \tag{12}$$

When a joint's orientation is updated, target orientations of its children also need to be updated, because the children have now been moved closer to the target. This also has the advantage of compensating for relative motion of the target: if the gaze shift is layered onto other body animation (e.g., walking or idle motion) or if the target itself is moving, continually recomputing joint target orientations ensures that the model accurately drives the joints toward the target. First we recompute fully aligned orientations $q_{\text{full},j}^{i+1}$ and then we compute the updated target orientation $q_{T,j}^{i+1}$ as follows:

$$q_{T,j}^{i+1} = slerp\left(q_{S,j}, q_{\text{full},j}^{i+1}, \frac{D_j^i}{\angle(q_{S,j}, q_{\text{full},j}^i)}\right) \tag{13}$$
$$D_j^i = \angle(q_{S,j}, q_{T,j}^i).$$

The parameter r_j^i must also be renormalized by multiplying it with D_j^i / D_j^{i+1}.

An additional consideration in our model is dynamic OMR. Guitton and Volle [1987] found that OMR gets narrower as head velocity increases. We simulate this property in our model and recompute the OMR on every frame using the following equation:

$$\text{OMR}^{i+1} = \text{OMR}^i \left(-\frac{1}{600} V_j^i + 1\right), \tag{14}$$

where V_j^i is the current head velocity, computed using Equation (10).

3.5. Ancillary Components of the Model

In addition to the main component of the model, which synthesizes gaze shifts as coordinated eye, head, and upper body movements, our gaze model also features ancillary components designed to increase the realism of the agent's behavior. The first of these components is a blink controller, which generates two types of behavior:

(1) *Gaze-evoked eye blinks* – These eye blinks occur during saccades, and their hypothesized purpose is to lubricate the cornea and protect the eye during movement [Evinger et al. 1994]. We generate gaze-evoked blinks probabilistically using the distribution proposed by Peters and Qureshi [2010].
(2) *Idle eye blinks* – These eye blinks occur when the eyes are not in a saccade. They are generated using an exponential distribution at the constant average rate of 20 blinks/*s* [Bentivoglio et al. 1997].

The second ancillary component is an idle gaze generator, an implementation of the Eyes Alive model presented by Lee et al. [2002]. This component generates random saccades based on a statistical model of idle gaze. Individual saccades are synthesized by our own model. The idle gaze behavior contributes to realism because it prevents the agent from continually staring at the same target; rather, the agent will periodically avert its gaze before looking back at the original target after a short time.

4. STUDY 1: MODEL VALIDATION

Development of the model was followed by an empirical evaluation of the model's communicative accuracy and naturalness. We conducted a study where we compared gaze shifts generated by our model against those generated by a state-of-the-art model [Peters and Qureshi 2010] as well as against gaze shifts displayed by human confederates. The study had two objectives. First, we needed to confirm that the model is able to accurately communicate attention direction. Second, we wished to confirm that gaze shifts generated by our neurophysiologically based model have comparable or better naturalness than those generated by a state-of-the-art procedural model.

Study results show that our model achieves the communicative accuracy of real human gaze and of the state-of-the-art model. Furthermore, results partially support our expectation that the model would achieve better naturalness ratings than the state-of-the-art model—gaze shifts generated by the model are rated as marginally more natural overall, while a subset of gaze shifts with full head alignment are rated as significantly more natural. In addition, we explored the effect that participant gender and the gender of the virtual character had on our measures, as gender is known to shape gaze perception in artificial agents [Mutlu et al. 2006].

4.1. Setup and Task

Participants observed a series of videos in which either an animated virtual character or a human confederate shifted his or her gaze toward one of 16 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 factors and facilitating the matching of animated and real-world conditions. Participants observed the agent from the perspective of a collaborator who would be seated across from the agent or the human confederate. The objects on the desk were separated into four groups and were distinguished by color and shape. Still images from the videos are shown in Figure 7.

The faces of the virtual characters used in the study were modeled using Singular Inversions FaceGen.[1] The faces and the eyes of both characters have approximately the same dimensions. The characters were scaled and positioned such that their size and location on the screen were as close as possible to those of human confederates.

[1]http://www.facegen.com.

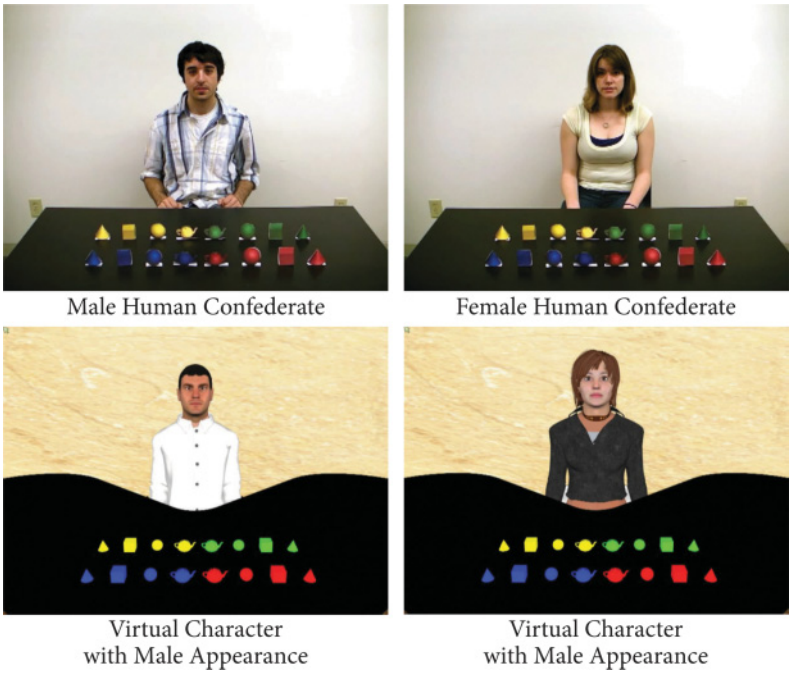

Fig. 7. Experimental conditions in Study 1.

4.2. Study Design

We conducted a 2 × 2 × 8 factorial experiment with a split-plot design. Our factors were:

(1) *Participant gender* – Two levels varying between participants
(2) *Virtual agent's gender* – Two levels varying between participants
(3) *Model type* – Eight levels varying within participants

The model type factor consisted of *six* comparison conditions for gaze shifts generated by the state-of-the-art head propensity model [Peters and Qureshi 2010] and those produced by our model, as well as *two* control conditions. The comparison conditions were created as follows. For each model, we manipulated the model parameters to create *three* distinct comparison conditions with different head alignment/propensity levels—0%, 50%, or 100%—adding up to a total of six such conditions in the experiment. This manipulation enabled us to determine how changes in the model parameters affected the communicative accuracy and perceived naturalness of the gaze shifts. In addition, the model type factor included two control conditions. In the first control condition, the virtual agent was replaced with a male or female human confederate, who presented gaze shifts toward the object on a desk in front of him or her. The second control condition involved a virtual agent maintaining gaze toward the participant without producing any gaze shifts.

4.3. 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, announced that they were about to look toward an object with 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 would guess which object the agent had gazed at, upon which they would fill out a questionnaire for subjective evaluation. Participants observed each gaze model generating gaze shifts toward all object types, colors, and positions. We randomized the order in which the participants observed the videos to minimize transfer effects. Each video was 10 seconds long, with the overall study lasting approximately 20 minutes.

4.4. Participants

Ninety-six participants (50 males and 46 females) took part in the study. The participants were recruited through Amazon.com's Mechanical Turk online marketplace, following crowdsourcing best practices to minimize the risk of abuse and to achieve a wide range of demographic representation [Kittur et al. 2008]. Participants received $2.50 for their participation.

4.5. Measures

The 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 human confederates or the virtual characters was directed. To measure perceived naturalness, we constructed a scale using four items that measured *naturalness*, *humanlikeness*, *lifelikeness*, and *realism*. These items were highly correlated (Cronbach's $\alpha = .94$). Participants rated each video for each item using a 7-point rating scale.

4.6. Results

We analyzed our data using a mixed-model analysis of variance (ANOVA).[2] Overall, model type had a significant effect on communicative accuracy, $F(7, 47.21) = 20.074, p < .0001$, and perceived naturalness, $F(7, 47.21) = 204.88, p < .0001$. In addition, for each measure we performed a total of *six* comparisons across conditions between our model and each control as well as our model against the baseline. We controlled family-wise error rate using the Dunn-Šidak test, with adjusted significant alpha of .00851 and marginal alpha of .0174. Results of these comparisons are presented later.

Communicative Accuracy – Comparisons across conditions found no significant differences in the communicative accuracy of the gaze shifts produced by our model, aggregated across all levels of head alignment, and those produced by human confederates. Similarly, no differences in accuracy were found between our model and the head propensity model, neither when aggregating across all levels of head alignment nor when comparing corresponding head alignment levels. The results suggest that the gaze shifts generated by our model are as accurate as those performed by human confederates and those generated by the state-of-the-art model. Unsurprisingly, our model achieves significantly higher accuracy than the agent control (virtual agent producing no gaze shifts), $F(1, 122.9) = 11.23, p = .0055$.

Perceived Naturalness – Comparisons across conditions showed that participants rated gaze shifts generated by our model, aggregated across all levels of head alignment, as marginally more natural than those generated by the head propensity model, aggregated across all levels of head propensity, $F(1, 2829) = 6.037, p = .014$. Pairwise comparisons across the two models with corresponding head alignment/propensity values showed that, at 100% alignment/propensity, participants rated gaze shifts produced by our model to be significantly more natural than those generated by the Peters model, $F(1, 2826) = 7.88, p = .005$. Furthermore, our model was rated as significantly less

[2]New analyses have been performed on the data from Study 1 and reported results differ from those in the original publication [Andrist et al. 2012b].

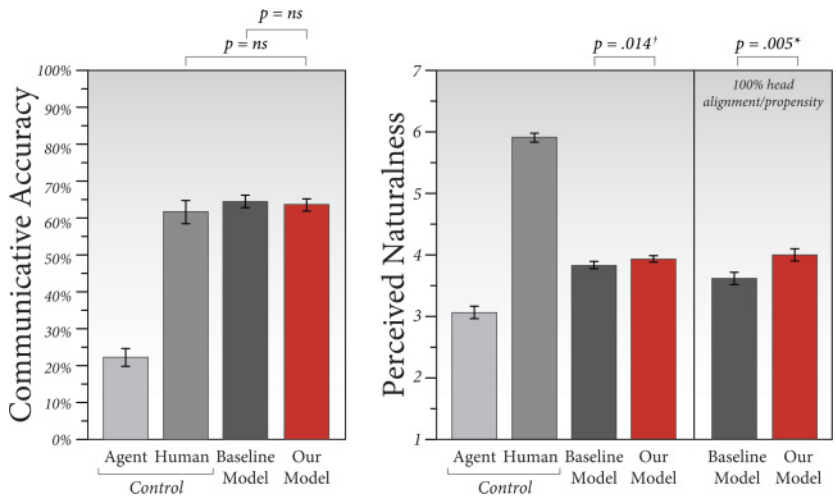

Fig. 8. Results of Study 1, aggregated across participant and agent gender. The baseline model refers to the head propensity model we used for comparison. From left to right: communicative accuracy, perceived naturalness aggregated across gaze shifts with different head alignments/propensity values, and perceived naturalness for the subset of gaze shifts with 100% head alignment/propensity.

natural than the human control, $F(1, 122.3) = 1108.247$, $p < .0001$, and significantly more natural than the agent control, $F(1, 122.9) = 193.21$, $p < .0001$.

Results from the communicative accuracy and perceived naturalness measures are illustrated in Figure 8.

Gender – Participants rated gaze shifts performed by the agent with female features as significantly more natural than those performed by the agent with male features, $F(1, 1958) = 17.17$, $p < .001$. On the other hand, communicative accuracy of the gaze shifts performed by the agent with male features was significantly higher than that of the shifts performed by the agent with female features, $F(1, 1958) = 4.85$, $p = .028$. Finally, we found a marginal interaction between participant gender and the gender features of the virtual character on the naturalness measure, $F(1, 1958) = 3.20$, $p = .074$. Figure 9 illustrates these results.

4.7. Discussion

Study results show that gaze shifts generated by our model communicate attention direction as accurately as do human gaze shifts and those generated by the state-of-the-art model. The mean accuracy for both models and the human control is below 65%, indicating that discriminating the correct gaze target using gaze as a visual cue is a difficult task for people. Comparable results are obtained by Bailly et al. [2010], who utilized a similar experimental protocol. In their study, participants inferred the correct object with 85% accuracy when aided by the virtual agent's gaze cues—the higher accuracy is likely due to lower spatial density of gaze targets in their setup.

Furthermore, the results of the current study show that our model produces marginally more natural gaze shifts than the state-of-the-art model across the range of the head alignment parameter, as well as significantly more natural gaze shifts at 100% head alignment. We attribute the difference in naturalness to the neurophysiological basis of our model, which should result in more realistic motions. However, a significant increase in naturalness is only achieved for gaze shifts at full head

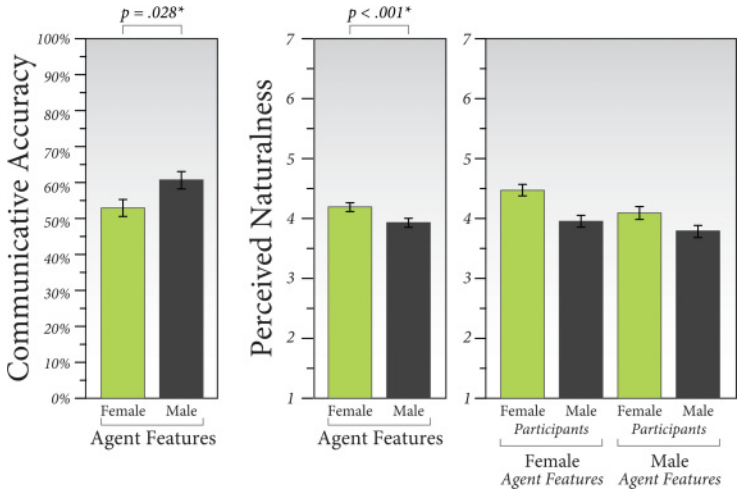

Fig. 9. Communicative accuracy and perceived naturalness across agents with male and female features.

alignment; we believe this is because those gaze shifts are dominated by head motion, which is slower and more visually prominent than the quick, subtle eye saccades.

Finally, we found significant differences in naturalness and communicative accuracy between the two conditions of agent gender, as well as marginal interaction between agent and participant gender. We believe these effects are caused by stylistic differences in character design. For example, the female character's gaze may be perceived as more natural because the character is more aesthetically appealing than the male character. Moreover, small variations in eye size and shape between the two characters could also affect people's perceptions of their gaze. We hypothesize that more consistent character design would reduce the effects of gender on experimental results.

5. STUDY 2: EYE–HEAD COORDINATION

We employed our gaze model in further studies with human participants to demonstrate how control over coordinated eye, head, and upper body movements enables agents to more effectively manage participants' attention and achieve better task performance and subjective perceptions. The study described in the current section involved a teaching task with a virtual agent. In this task, the agent gave a lecture on Chinese history while using a virtual map of China as instructional material. During the lecture, the agent distributed its gaze between the participant and a virtual map. We controlled eye–head coordination using the head alignment parameter to produce distinct patterns of gaze behavior—*affiliative gaze*, where the agent kept its head turned toward the participant and gazed at the map "out of a corner of the eye" (thus emphasizing eye contact, Figure 10, top row), and *referential gaze*, where the agent had its head turned toward the map and only glanced at the participant (emphasizing information shown on the map, Figure 10, bottom row). While the agent gazed at the same targets in the same sequence in both cases, we expected that the agent using affiliative gaze would elicit stronger feelings of affiliation with the participant and therefore would be perceived more positively, while referential gaze would enable the participant to create stronger associations between verbal utterances and map information, resulting in better recall of lecture information.

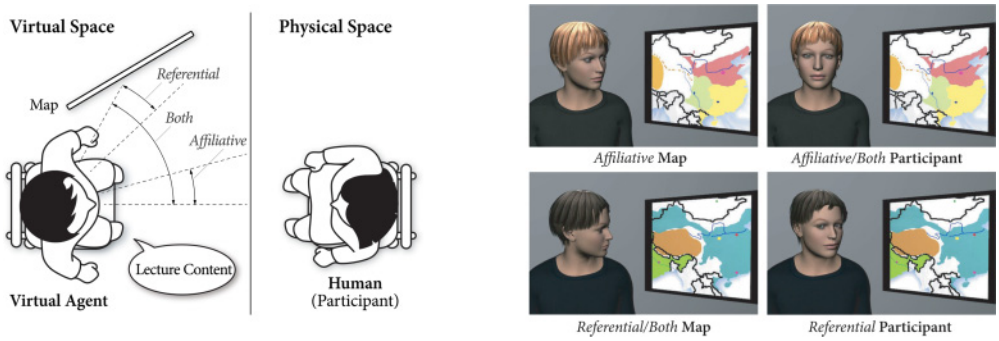

Fig. 10. Left: Diagram showing the range of the agent's head movements for different gaze conditions. Right: Experimental conditions in Study 2. From top left in clockwise order: *affiliative*, looking at map; *affiliative*, looking at participant; *referential*, looking at map; *referential*, looking at participant.

5.1. Hypotheses

We designed an experiment to test the following three hypotheses:

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

The literature strongly supports the first hypothesis—it has been shown that students learn better when they are gazed at by their teachers [Burgoon et al. 1986; Fry and Smith 1975; Sherwood 1987; Otteson and Otteson 1980].

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

This hypothesis is supported by the fact that people are perceived as being more intelligent, more trustworthy, and more friendly when they make more direct eye contact [Argyle and Cook 1976; Fullwood and Doherty-Sneddon 2006]. We wish to extend this work to show that head alignment makes an impact on the positive subjective effects of mutual gaze.

Hypothesis 3 – Interacting with an agent that uses more referential gaze will result in better recall performance. This will especially be true when the information to be recalled relies on building associations with objects in the environment (in our case, geographic locations marked on the map).

Referential gaze helps build associations between information and objects in the environment. Eye gaze is an important nonverbal cue in establishing common ground [Clark and Brennan 1991]; for example, when a speaker makes verbal references to information in the environment, his or her gaze might shift between looking at the information and looking at the listener to ensure his or her continued attention and understanding [Nakano et al. 2003]. In communication between a human and a virtual agent, it has been shown [Liu et al. 2011] that perceiving the object of the agent's gaze reduces the time and verbal communication needed for grounding references. We believe that head alignment has the ability to strengthen or weaken this effect, as it plays a substantial role in shifting a viewer's visual attention [Hietanen 1999]. Full alignment of the agent's head with the referenced object should serve as a stronger referential gaze cue than when the head is not aligned.

5.2. 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 online student job forums.

5.3. Study Design

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

(1) *Audio* – The agent, who is visible to the participant, is introduced briefly. The lights in the virtual scene are extinguished for the duration of the lecture except for a spotlight on the map; thus, the agent is in darkness and only the map can be seen. At the end of the lecture, the scene lights turn back on and the agent gives the participant instructions for the subjective evaluation and lecture quiz.
(2) *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, its head is fully aligned with the participant. When the agent shifts its eye gaze to refer to something on the map, the head turns as little as possible to the map, so the agent glances sideways at the map while keeping its head turned toward the participant. Some head movement toward the map is necessary due to restricted OMR; otherwise, the agent's eyes would be unable to reach the map.
(3) *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 so that the agent glances sideways at the participant while keeping the head aligned toward the map. Some head movement toward the participant is necessary due to restricted OMR.
(4) *Both* – This is a cross between the previous two conditions. When gazing at the participant, the agent keeps its head fully aligned with the participant. When gazing at the map, the agent aligns its head fully with the map.

Figure 10 illustrates the differences in head motion between three of the conditions: *affiliative*, *referential*, and *both*. Each image shows what the agent looked like during gaze fixations as it gazed at the map or the participant, respectively. As the agent performed gaze shifts to and from the map, its eyes and head moved between these terminal poses. We note that, in each condition, the agent always started out in the same pose—gazing straight ahead at the participant. The agent would maintain that pose during the introductory part of the lecture (before the first gaze shift) and revert to it during the instruction phase at the end of the lecture. In the *audio* condition, the agent is only ever seen in the starting pose; the scene is unlit during the lecture, so the agent displays no gaze shifts.

Each participant viewed four lectures delivered by four different virtual agents, each of which utilized a different gaze condition; thus, every participant was exposed to all four gaze conditions. All agents, two of which can be seen in Figure 10, were specifically designed to look and sound androgynous 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. 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 approximately 3 minutes in length.

During each lecture, the agent made 11 gaze shifts to the map, each followed shortly by a gaze shift back to the participant. The agent fixated its gaze on different targets

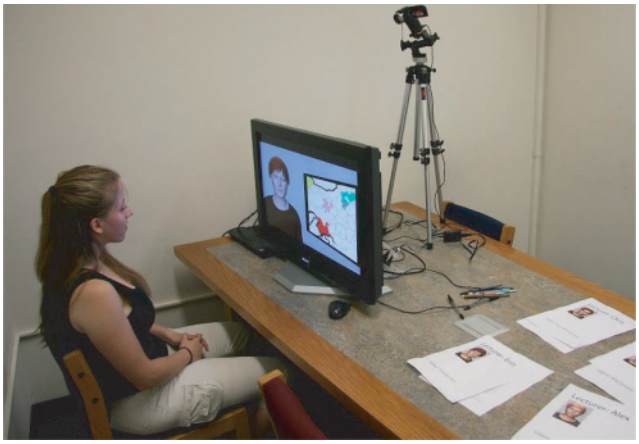

Fig. 11. Experimental setup in Study 2.

on the map as they were mentioned, for example, while giving facts about a Chinese city indicated on the map. We specified all the gaze shifts manually by annotating the lectures with gaze cues in advance and then triggering those gaze shifts as the lecture proceeded. Griffin [2001] and Meyer et al. [1998], which indicate that a deictic gaze shift should occur 800 to 1,000*ms* before the object being gazed at is referenced in speech, guided our gaze shift timings.

5.4. Implementation

To facilitate the creation of experimental scenarios, we built a virtual agent framework within the Unity game engine.[3] Various agent behaviors (gaze, random eye movement, blinking, etc.) are implemented as reusable Unity C# scripts and are robust enough to scale to characters with varying skeletal structures and morphologies.

We also devised a robust pipeline that allows us to quickly produce sets of character models and import them into Unity. Our gender-neutral agent models were created using DAZ Studio[4] by parametrically modifying a base figure. The characters were exported into Autodesk 3ds Max for further editing. We then produced agent voice-overs, adapted their pitch using Audacity to make them sound androgynous, and generated lip-sync animations using the FaceFX[5] 3ds Max plugin. Finally, we exported the characters into Unity and extended them with our behavior scripts, yielding a set of fully functional virtual agents.

5.5. Procedure

The experiment was conducted in a closed study room with no outside distractions. Participants were brought into the room and asked to sit at a table with a computer monitor and mouse. The monitor was a 32-inch flat-panel display, allowing the virtual agent representation (only the head and shoulders) to be near life size. This setup can be seen in Figure 11. The participant gave informed consent after the experimenter briefly described the study tasks to the participant. 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.

[3]http://www.unity3d.com.
[4]http://www.daz3d.com.
[5]http://www.facefx.com.

The experimenter told participants that after he left the room, they could press the start button on the screen. Upon pressing the start button, the first lecturer (randomly chosen by the software) introduced itself and gave 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 allowed the subjective evaluation to double as a distracter task, strengthening any subsequent recall measures. All participants were monitored via closed-circuit camera by the experimenter to ensure that these instructions were followed. No participants were observed to neglect the instructions.

After filling out the subjective evaluation and quiz, the participant went back to the monitor and clicked the on-screen button to begin the next lecture. This process was repeated until all four lectures were 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 paid the participant. The total experiment took approximately 30 minutes, and participants were paid $5 for their participation.

5.6. Measures

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

The objective measurement of recall took the form of quizzes taken by all participants following each lecture. Each quiz contained 10 short-answer questions, which 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 is "In what year did the Jin dynasty overtake control of China?" This information 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 such question was "Which of the *Three Kingdoms* dynasties extended farthest south?" This question is answerable only after study of the map. The third category, which included the remaining three questions of the quiz, relied on the building of 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 10-question quiz.

The subjective measurements were split into six broad indicators, and each question within the indicators took the form of a 7-point rating scale.

(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 participant rapport with the agent. Questions asked included, for example, how well the participant felt he or she connected with the agent and how willingly he or she would disclose personal information to the agent following the lecture (Cronbach's $\alpha = .84$).
(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 the participant perceived the agent to be at conveying lecture material (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 – Our postlecture questionnaires also included questions to check whether our manipulations were effective. First, we wanted to test whether the agents we created were actually being perceived as androgynous. Two questions in the form of 7-point rating scales, anchored by *very feminine* (value = 7) and *very masculine* (value = 1), were asked. One referred to the agent's *appearance* and the other to the agent's *voice*.

Second, we wished to check that the *gaze type* manipulations between the visible agent conditions were noticed by participants. 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.

5.7. Results

Data analysis 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. The analysis started with the manipulation checks for the perceived gender of the agent and the *gaze type* manipulations. First, we found that the agents were rated on average as being mostly androgynous (appearance: M = 4.88, SD = 1.21, voice: M = 4.51, SD = 1.58). Second, we found that participants felt more attended to in the *affiliative* gaze condition versus the *referential* condition, $F(1, 69) = 12.53$, $p < .001$. Participants also felt more attended to in the *both* condition than in the *referential*, $F(1, 69) = 4.37$, $p = .040$. Finally, participants felt that the agent attended to the map more in the *referential* versus *affiliative* condition, $F(1, 69) = 7.75$, $p = .007$. 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, $F(1, 69) = 4.75$, $p = .033$.

We next analyzed the objective results in the form of recall quiz scores (Figure 12). 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 10 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* also showed no significant difference, $F(1, 69) = 0.59$, $p = .45$.

Finally, we analyzed the subjective measures. 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

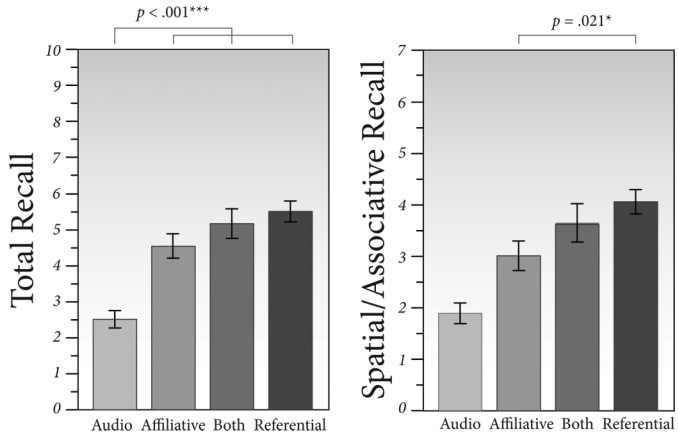

Fig. 12. Results of objective measure (recall measured by postlecture quiz). On the left is the total quiz performance; on the right is the quiz performance when only considering a subset of the questions: those dealing with spatial information and building associations.

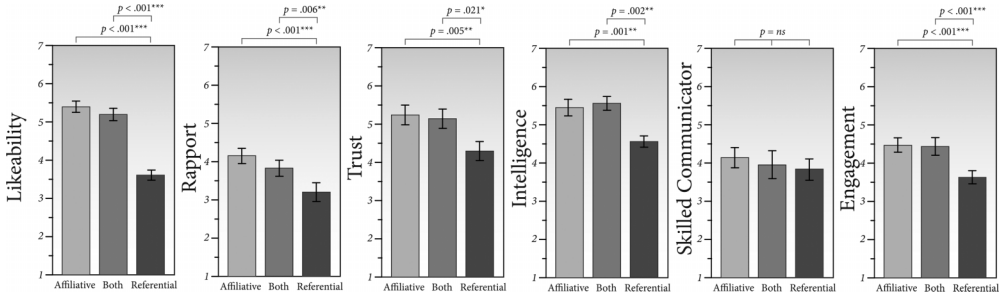

Fig. 13. Results of subjective evaluations (likeability, rapport, trust, intelligence, skilled communicator, and engagement) based on gaze condition.

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 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 between conditions, but the *engagement* scale showed that the *referential* condition 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 13.

5.8. Discussion

This study was conducted to demonstrate how changes in eye–head coordination during gaze shifts can have significant effects on the interaction between an agent and a human, resulting in improved learning and positive feelings of affiliation. To show this, we manipulated the *head alignment* parameter on various virtual teaching agents. By doing so, we modified the head movement in the agent's gaze shifts from the participant to an object in the environment (map of China) and back again, without affecting either the number or the sequence of the gaze shifts. We showed that by manipulating just this one gaze parameter, we could achieve very different task outcomes, measured by information recall and the participant's evaluation of the agent.

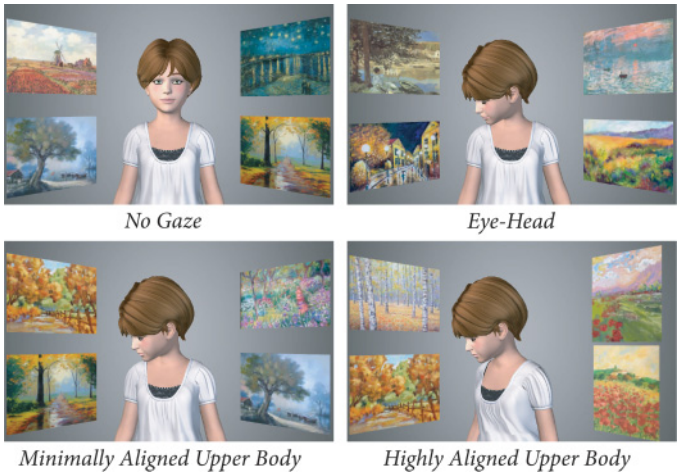

Fig. 14. Experimental conditions in Study 3.

Confirming our first hypothesis, the mere presence of an agent resulted in better recall performance than audio alone, irrespective of the gaze condition. For five of the six subjective scales, the *affiliative* and *both* gaze conditions achieved better ratings than the *referential* condition. This leads us to accept the second hypothesis, showing that human listeners prefer the agent to fully align its head with the participant while speaking rather than to look out of the corners of its eyes. 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 regardless of the gaze condition. We also have strong support for the third hypothesis, since 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. Since both conditions had the same number and sequence of gaze shifts, 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, suggesting that not all methods of gazing are equal and that changes in eye–head coordination can create significant differences in interaction outcome.

6. STUDY 3: UPPER BODY COORDINATION

The previous study showed that by manipulating a property of eye–head coordination in the agent's gaze shifts—the head alignment parameter—we can control how participants redirect their attention and produce positive effects in their interactions with the agent. The goal of the next study was to show that interaction outcomes can be similarly modified by manipulating an analogous property of upper body coordination in gaze—specifically, the trunk alignment parameter. In the third study, a virtual agent gazes at a series of paintings. We use the trunk alignment parameter to control how much the agent's upper body participates in these gaze shifts (Figure 14). Our expectation was that when the agent turned its upper body more toward a painting, this would produce a greater shift in attention in a participant than merely turning the agent's eyes and head, so the participants would attribute to the agent a greater interest in that painting.

6.1. Hypotheses

Our hypotheses for the study were the following:

Hypothesis 1 – Participants will perceive the agent as more interested in an object when the agent gazes at the object than when it ignores the object. This hypothesis is strongly supported in the literature, as the role of eye gaze in attention cueing and establishing joint attention has been extensively studied [Frischen et al. 2007].

Hypothesis 2 – Participants will perceive the agent as more interested in an object when the agent rotates its upper body toward the object, rather than rotating just its eyes and head.

Hypothesis 3 – Participants will perceive the agent as more interested in an object when the agent rotates its upper body toward the object with high trunk alignment rather than with minimal trunk alignment.

The latter two hypotheses have support in prior work in the social sciences, which suggests a link between body orientation and human attention cueing mechanisms [Hietanen 2002; Pomianowska et al. 2011; Kendon 1973; Schegloff 1998]. On that basis, we believe that increased amount of upper body reorientation in gaze shifts suggests a greater shift in attention.

6.2. Participants

We recruited 15 participants (eight males, seven females) from a campus population using fliers, online job postings, and in-person recruitment. The study took 10 minutes and participants were paid $2 for their participation.

6.3. Study Design

The study followed a within-participants design with a single factor with four levels. Experimental conditions were as follows:

(1) *No gaze* – The agent does not look toward the painting.
(2) *Eye–head* – The agent gazes at the painting by turning its eyes and head toward the painting, without engaging the trunk.
(3) *Minimally aligned upper body* – The agent also turns its trunk toward the painting by a minimal amount (model parameter α_T set to 0).
(4) *Highly aligned upper body* – The agent turns its trunk toward the painting by a large amount (model parameter α_T set to 0.3).

Figure 14 depicts each of the four conditions.

The participant was asked to watch the agent perform sequences of gaze shifts toward four virtual paintings placed in the scene. The four paintings were static and fully visible to the participant at all times. For each set of paintings, the agent performed three gaze shifts—eye–head gaze, gaze with minimal upper body alignment, and gaze with high upper body alignment. In each case, the agent gazed at a different painting. In any set of paintings, the agent would gaze at three out of the four paintings *once* and would never gaze at the fourth painting. The placement of paintings and assignment of gaze conditions were randomized. After watching the gaze shifts, the participant was asked to rate the agent's interest in each painting. The figure on the left in Figure 15 shows the task interface, while the figure on the right shows the physical setup.

The task was implemented using the virtual agent framework we developed for Study 2 (Section 5.4).

6.4. Procedure

After recruitment, each participant read and signed a consent form. The participant then performed eight trials of the experimental task. The first three trials were used

Fig. 15. Experimental task in Study 3. Left: Task interface. Right: An experimenter demonstrating the task.

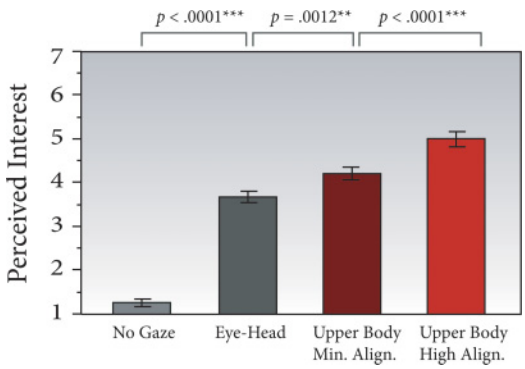

Fig. 16. Results for perceived interest ratings for each type of gaze shift. Results for gaze shifts with upper body movement are shown in red.

to acclimate the participant to the task, and data from these trials were not used. At the end of each trial, the participant rated the agent's interest in each painting using rating scales that appeared under each painting on the screen (Figure 15, left). After submitting their ratings, a new set of paintings and gaze shifts were presented to the participant. Task duration was approximately 5 minutes. At the end of the task, each participant filled out a demographic questionnaire.

6.5. Measures

The study used one subjective measure—*perceived interest*—which was measured using a single 7-point rating scale item (1 = "Uninterested," 7 = "Very Interested").

6.6. Results

Figure 16 shows the study results. Mean interest ratings were 1.25 (SD = 0.19) in the *no-gaze* condition, 3.67 (SD = 0.19) in the *eye–head* condition, 4.20 (SD = 0.19) in the *minimally aligned upper body* condition, and 5.00 (SD = 0.19) in the *highly aligned upper body* condition. One-way within-participants ANOVA found significant differences between means, $F(3, 282) = 195.03$, $p < .0001$.

A priori comparisons showed that the agent was perceived as significantly more interested in the *eye–head* condition than in the *no-gaze* condition, $F(1, 282) = 218.12$, $p < .0001$, providing support for the first hypothesis. Furthermore, the agent also was perceived as significantly more interested in the *minimally aligned upper body* condition than in the *eye–head* condition, $F(1, 282) = 10.65$, $p = .0012$, leading us to accept the

second hypothesis. Finally, the agent was perceived as significantly more interested in the *highly aligned upper body* condition than in the *minimally aligned upper body* condition, $F(1, 282) = 23.97$, $p < .0001$, which supports the third hypothesis.

6.7. Discussion

The purpose of this study was to show that changes in upper body orientation in gaze shifts produced by our gaze model strengthened the shift in attention signaled by the gaze shifts and led to increases in the perception of the agent's interest in objects in the environment. The results of the study support all of our hypotheses. As expected, the agent was perceived as being more interested in an object in a gaze condition than a no-gaze condition; moreover, engagement of the upper body during gaze shifts communicated interest more strongly than just eye and head movements. Trunk rotation amplitude in the *minimally aligned upper body* condition never exceeded 6°, yet even such a small amount of movement yielded a significant increase in the perceived interest score, suggesting that humans are highly sensitive to differences in body orientation. High trunk alignment yielded changes in trunk orientation in excess of 26° and led to a further increase in interest scores, suggesting a link between the agent's upper body orientation, controlled using the trunk alignment parameter, and the attention signaling properties of gaze.

7. GENERAL DISCUSSION

7.1. Gaze Model

The gaze model presented in the current work can synthesize gaze shifts that incorporate parametrically controllable eye, head, and upper body movements, which have been shown to have the communicative accuracy of a real human gaze (Section 4). As such, the model is a suitable building block for gaze behaviors of embodied conversational agents, which are designed by composing sequences of gaze shifts synthesized by the model. Designers can compose these behaviors manually to accompany a predefined interaction, as we did in Studies 2 and 3. These behaviors could also be generated automatically by a high-level model, which would specify gaze targets and gaze shift timings, while the synthesis of actual gaze shifts would be accomplished by our low-level model. For example, Andrist et al. [2013] employed the model as part of their gaze aversions controller for ECAs, which triggers conversational gaze aversions based on a set of probability distributions.

The model as presented in the current article is only applicable to agents with realistic, humanlike designs. Applying the gaze shifts synthesized by the model to stylized, cartoonlike embodiments would result in animation artifacts due to the exaggerated and unrealistic proportions of such characters. We have since developed a set of motion adaptation techniques for gaze that allow the model to work even on stylized embodiments [Pejsa et al. 2013]. As part of that endeavor, we conducted a validation study of the extended model similar to the one described in Section 4, which showed that gaze adapted to stylized agents has the same communicative accuracy and perceived naturalness as the gaze of realistically proportioned, humanlike agents.

Although the presented model offers a comprehensive solution for coordination of the eyes, head, and upper body during gaze shifts, there is room for further extensions. Specifically, the model could be extended to also animate pelvis orientation, as well as orientation of the whole body through repositioning of the feet. These extensions are necessary to have a truly comprehensive model of attention shifts, as we infer the direction of attention of others by observing the positioning of their entire body [Langton et al. 2000; Hietanen 2002; Pomianowska et al. 2011]. Extending our model to incorporate whole-body orientation would require the introduction of new alignment

parameters for controlling whole-body coordination, as well as improving the animation techniques employed in the model. One shortcoming of the current model is that upper body movements lack naturalness compared to eye and head movements. This is mainly due to the complexity of the human body, which makes purely procedural animation approaches inadequate for animating body orientation in a natural and humanlike way. One possible solution is to employ a synthesis-by-example approach, which would synthesize novel gaze shift motions from example motions obtained using motion capture [Pejsa and Pandzic 2010]. The challenge of synthesis-by-example approaches is in achieving sufficient control over motion, which can be difficult in that the range of synthesizeable motions is limited by the motion capture examples available.

7.2. Studies of Gaze for ECAs

Having built our model and validated it in Study 1, we employed it in two studies where we manipulated the model's alignment parameters to manage the attention of the participants and thus control subjective and objective task outcomes. In Study 2, the model's head alignment parameter was manipulated to produce several distinct patterns of gaze behavior. Affiliative gaze emphasized eye contact and triggered feelings of affiliation between the participant and agent, leading to higher subjective evaluations of the agent. Referential gaze emphasized information in the environment and caused the participant to form stronger associations between the information and verbal references, leading to better recall of information from the interaction. These results inform the designer as to how he or she can manipulate the agent's gaze behavior using a model parameter—head alignment—to manage a person's attention over the course of an interaction. If the designer judges that it is more important for a person to build a strong relationship with the agent, he or she can employ higher head alignment values for gaze shifts that establish eye contact. On the other hand, if it is more important for a person to pay attention to objects in the environment, the designer can use higher head alignment values for gaze shifts toward these objects.

In Study 3, we manipulated the model's trunk alignment parameter to control the amount of upper body reorientation in gaze shifts toward objects in the environment. The agent would look at objects either by directing its eyes and head toward the objects while keeping its upper body oriented to the participant or by turning the eyes, head, and upper body toward the objects by a minimal or large amount. In the latter case, the agent was perceived as expressing more interest in the objects in question, indicating that a change in upper body orientation acts as a significant attention cue. What is more, even a very small change in body orientation led to significant increases in perceived interest, while for large changes in body orientation those increases were greater. These findings suggest that the designer can manipulate trunk alignment to continuously vary the perception of the agent's interest, analogous to how head alignment manipulation can be used to control information recall and affiliation. Furthermore, they suggest that attention cueing effects could be strengthened further by introducing reorientation of the agent's entire body.

The significance of our studies is in that they show how ECA's gaze can trigger beneficial social and cognitive processes and lead to desired interaction outcomes, and also in that they inform designers as to how they can control these outcomes in a continuous way by manipulating specific model parameters—head alignment and trunk alignment. Furthermore, these manipulations only affect the spatial coordination properties of individual gaze shifts, while the sequence and timings of gaze shifts remain unchanged. The studies therefore illustrate how manipulating these underexplored properties of gaze can significantly enhance an agent's communicative function.

Our work forms a strong basis for continued research in gaze for ECAs. In particular, the presented model, with its demonstrated effectiveness in eliciting strong social and

cognitive effects in humans interacting with the agent, can serve as a building block of gaze mechanisms that in turn facilitate more complex conversational processes, such as floor management and establishment of conversational roles. The previously mentioned gaze aversion model by Andrist et al. [2013] employed the current gaze shift model to synthesize conversational gaze aversions in face-to-face conversations. This work showed that an agent can use gaze aversions in an interview-style interaction to achieve smoother turn taking, higher disclosure, and improved subjective perceptions of the agent. We believe that even more powerful conversational mechanisms can be constructed through incorporation of body reorientation, which serves as a powerful attention cue in the context of multiparty interaction. Prior research in the social sciences has shown that people reorient their bodies to establish conversational roles and reconfigure conversational formation [Kendon 1973, 2010]. Research in human–robot interaction [Kuzuoka et al. 2010] has shown that humanlike robots can achieve the same. This ability is particularly important in a dynamic setting, where parties can join or leave the conversation at any time and conversational roles fluctuate constantly—for example, agents used in the service industry, educational and guide agents, and avatars in online games and virtual worlds. We believe the current work on head and body orientation in gaze shifts presents a solid basis for exploration of issues in multiparty interactions with an embodied conversational agent.

8. CONCLUSIONS

Humans employ gaze cues to communicate their attention direction, initiate eye contact, and establish reference and joint attention when they interact. By doing so, they trigger important social and cognitive processes, such as positive feelings of affiliation and improved learning of information from speech and the environment. The fundamental unit of these behaviors is the gaze shift, which consists of coordinated movements of the eyes, head, and upper body. Interpretation of gaze in communication depends not only on eye movements (saccades) but also on movements of the head and body that occur in concert with saccades, as humans infer attention direction of others by integrating information about the orientation of their eyes, head, and body.

In the current work, we built a computational model for synthesis of ECA's gaze shifts that enables the designer to parametrically control how the eyes, head, and upper body coordinate during the gaze shift. The model exposes head and trunk alignment parameters, which specify how much the head and trunk will turn during the gaze shift, respectively. A validation study with human participants showed that the model generates gaze that conveys attention direction as accurately as real human gaze and achieves improvements in naturalness over a state-of-the-art model. Moreover, we conducted further two studies with human participants to demonstrate that control over the coordination parameters in agents' gaze shifts affords better management of the participants' attention, which can be used to achieve desired social and cognitive effects in participants interacting with the agent.

The value of our work lies not only in the model of gaze shifts that offers rich parametric control and enables improvements in agents' task performance and subjective perceptions but also in that we specifically demonstrate how eye–head and eye–body coordination properties of gaze shifts can be manipulated in a principled and continuous way to manage the attention of humans interacting with the agent in order to achieve desired interaction outcomes. A gaze model enabling manipulation of these properties can serve as a building block of larger mechanisms of social gaze, such as referential gaze for conveying information embedded in the environment, mutual gaze for building affiliative relationships between the agent and people, and floor management gaze in two-party and multiparty conversations.

REFERENCES

Y. Alkan, B. B. Biswal, and T. L. Alvarez. 2011. Differentiation between vergence and saccadic functional activity within the human frontal eye fields and midbrain revealed through fMRI. *PLoS One* 6, 11 (Nov. 2011).

C. André-Deshays, A. Berthoz, and M. Revel. 1988. Eye-head coupling in humans. *Experimental Brain Research* 69, 2 (1988), 399–406.

S. Andrist, B. Mutlu, and M. Gleicher. 2013. Conversational gaze aversion for virtual agents. In *Proceedings of 13th International Conference on Intelligent Virtual Agents (IVA'13)*, R. Aylett, B. Krenn, C. Pelachaud, and H. Shimodaira (Eds.). Vol. 8108. Springer, 249–262.

S. Andrist, T. Pejsa, B. Mutlu, and M. Gleicher. 2012a. Designing effective gaze mechanisms for virtual agents. In *Proceedings of ACM/CHI 2012 Conference on Human Factors in Computing Systems*. ACM, 705–714.

S. Andrist, T. Pejsa, B. Mutlu, and M. Gleicher. 2012b. A head-eye coordination model for animating gaze shifts of virtual characters. In *Proceedings of 4th ICMI Workshop on Eye Gaze in Intelligent Human Machine Interaction (Gaze-In'12)*. ACM, 4:1–4:6.

M. Argyle and M. Cook. 1976. *Gaze and Mutual Gaze*. Cambridge University Press.

A. T. Bahill, M. R. Clark, and L. Stark. 1975. The main sequence, a tool for studying human eye movements. *Mathematical Biosciences* 24, 34 (1975), 191–204.

G. Bailly, S. Raidt, and F. Elisei. 2010. Gaze, conversational agents and face-to-face communication. *Speech Communication* 52, 6 (June 2010), 598–612.

G. R. Barnes. 1979. Vestibulo-ocular function during co-ordinated head and eye movements to acquire visual targets. *Journal of Physiology* 287, 1 (1979), 127–147.

S. A. Beebe. 1976. *Effects of Eye Contact, Posture and Vocal Inflection upon Credibility and Comprehension*. Ph.D. Dissertation. University of Missouri-Columbia.

A. R. Bentivoglio, S. B. Bressman, E. Cassetta, D. Carretta, P. Tonali, and A. Albanese. 1997. Analysis of blink rate patterns in normal subjects. *Movement Disorders: Official Journal of the Movement Disorder Society* 12, 6 (1997), 1028–1034.

J. K. Burgoon, D. A. Coker, and R. A. Coker. 1986. Communicative effects of gaze behavior. *Human Communication Research* 12, 4 (1986), 495–524.

A. Cafaro, R. Gaito, and H. Vilhjálmsson. 2009. Animating idle gaze in public places. In *Proceedings of 9th International Conference on Intelligent Virtual Agents*. Springer, 250–256.

J. Cassell, O. Torres, and S. Prevost. 1999. Turn taking vs. discourse structure: How best to model multimodal conversation. *Machine Conversations*, Y. Wilks (Ed.), Kluwer, 143–154.

H. H. Clark and S. A. Brennan. 1991. Grounding in communication. In *Perspectives on Socially Shared Cognition*, L. B. Resnick, J. M. Levine, and S. D. Teasley (Eds.). APA Books.

B. D'Entremont, A. Yazbeck, A. Morgan, and S. MacAulay. 2007. Early gaze-following and the understanding of others. In *Gaze Following: Its Development and Significance*, R. Flomm, K. Lee, and D. Muir (Eds.), Lawrence Erlbaum Associates, 77–94.

C. Evinger, K. A. Manning, J. J. Pellegrini, M. A. Basso, A. S. Powers, and P. A. Sibony. 1994. Not looking while leaping: The linkage of blinking and saccadic gaze shifts. *Experimental Brain Research* 100, 2 (1994), 337–344.

J. Fox and J. N. Bailenson. 2009. 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.

E. G. Freedman and D. L. Sparks. 2000. Coordination of the eyes and head: Movement kinematics. *Experimental Brain Research* 131, 1 (2000), 22–32.

A. Frischen, A. P. Bayliss, and S. P. Tipper. 2007. Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin* 133, 4 (2007), 694–724.

R. Fry and G. F. Smith. 1975. The effects of feedback and eye contact on performance of a digit-coding task. *Journal of Social Psychology* 96, 1 (1975), 145–146.

J. H. Fuller. 1992. Head movement propensity. *Experimental Brain Research* 92, 1 (1992), 152–164.

C. Fullwood and G. Doherty-Sneddon. 2006. Effect of gazing at the camera during a video link on recall. *Applied Ergonomics* 37, 2 (2006), 167–175.

M. Garau, M. Slater, S. Bee, and M. A. Sasse. 2001. The impact of eye gaze on communication using humanoid avatars. In *Proceedings of ACM/CHI 2001 Conference on Human factors in Computing Systems*. ACM, 309–316.

G. N. Goldberg, C. A. Kiesler, and B. E. Collins. 1969. Visual behavior and face-to-face distance during interaction. *Sociometry* 32, 1 (1969), 43–53.

J. E. Goldring, M. C. Dorris, B. D. Corneil, P. A. Ballantyne, and D. R. Munoz. 1996. Combined eye-head gaze shifts to visual and auditory targets in humans. *Experimental Brain Research* 111, 1 (1996), 68–78.

H. H. Goossens and A. J. V. Opstal. 1997. Human eye-head coordination in two dimensions under different sensorimotor conditions. *Experimental Brain Research* 114, 3 (1997), 542–560.

Z. M. Griffin. 2001. Gaze durations during speech reflect word selection and phonological encoding. *Cognition* 82, 1 (2001), B1–B14.

H. Grillon and D. Thalmann. 2009. Simulating gaze attention behaviors for crowds. *Computer Animation and Virtual Worlds* 20, 23 (June 2009), 111–119.

E. Gu and N. I. Badler. 2006. Visual attention and eye gaze during multiparty conversations with distractions. In *Proceedings of 6th International Conference on Intelligent Virtual Agents (IVA'06)*. Springer-Verlag, 193–204.

D. Guitton and M. Volle. 1987. Gaze control in humans: Eye-head coordination during orienting movements to targets within and beyond the oculomotor range. *Journal of Neurophysiology* 58, 3 (1987), 427–459.

M. J. Harris and R. Rosenthal. 2005. No more teachers' dirty looks: Effects of teacher nonverbal behavior on student outcomes. *Applications of Nonverbal Communication*, R. E. Riggio and R. S. Feldman (Eds.), Lawrence Erlbaum Mahwah, 157–192.

R. M. Heck. 2007. *Automated Authoring of Quality Human Motion for Interactive Environments*. Ph.D. Dissertation. University of Wisconsin–Madison.

D. Heylen, I. Van Es, E. Van Dijk, A. Nijholt, and B. Van Dijk. 2002. Experimenting with the gaze of a conversational agent. In *Proceedings of International CLASS Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems*. Kluwer Academic Publishers, 93–100.

J. K. Hietanen. 1999. Does your gaze direction and head orientation shift my visual attention? *Neuroreport* 10, 16 (1999), 3443.

J. K. Hietanen. 2002. Social attention orienting integrates visual information from head and body orientation. *Psychological Research* 66, 3 (2002), 174–179.

M. A. Hollands, N. V. Ziavra, and A. M. Bronstein. 2004. A new paradigm to investigate the roles of head and eye movements in the coordination of whole-body movements. *Experimental Brain Research* 154, 2 (2004), 261–266.

D. H. Kelley and J. Gorham. 1988. Effects of immediacy on recall of information. *Communication Education* 37, 3 (1988).

A. Kendon. 1973. The role of visible behaviour in the organization of social interaction. In *Social Communication and Movement*, M. V. Cranach and I. Vine (Eds.), Academic Press, 29–74.

A. Kendon. 2010. Spacing and orientation in co-present interaction. In *Development of Multimodal Interfaces: Active Listening and Synchrony*. Lecture Notes in Computer Science, Vol. 5967. Springer, Berlin, 1–15.

S. C. Khullar and N. I. Badler. 2001. Where to look? Automating attending behaviors of virtual human characters. *Autonomous Agents and Multi-Agent Systems* 4, 1 (2001), 9–23.

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

A. Kittur, E. H. Chi, and B. Suh. 2008. Crowdsourcing user studies with Mechanical Turk. In *Proceedings of ACM/CHI 2008 Conference on Human factors in Computing Systems (CHI'08)*. ACM, 453–456.

H. Kuzuoka, Y. Suzuki, J. Yamashita, and K. Yamazaki. 2010. Reconfiguring spatial formation arrangement by robot body orientation. In *Proceedings of 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI'10)*. 285–292.

B. J. Lance and S. C. Marsella. 2010. The expressive gaze model: Using gaze to express emotion. *IEEE Computer Graphics & Applications* 30, 4 (2010), 62–73.

S. R. H. Langton, R. J. Watt, and V. Bruce. 2000. Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences* 4, 2 (2000), 50–59.

J. Lee, S. Marsella, D. Traum, J. Gratch, and B. Lance. 2007. The rickel gaze model: A window on the mind of a virtual human. In *Proceedings of 7th International Conference on Intelligent Virtual Agents (IVA'07)*. Springer, 296–303.

S. P. Lee, J. B. Badler, and N. I. Badler. 2002. Eyes alive. *ACM Transactions on Graphics* 21, 637–644.

Z. Li, X. Mao, and L. Liu. 2009. Providing expressive eye movement to virtual agents. In *Proceedings of the 2009 International Conference on Multimodal Interfaces (ICMI-MLMI'09)*. 241–244.

C. Liu, D. L. Kay, and J. Y. Chai. 2011. Awareness of partner's eye gaze in situated referential grounding: An empirical study. In *Proceedings of the 2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction*.

M. M. Lusk and R. K. Atkinson. 2007. Animated pedagogical agents: Does their degree of embodiment impact learning from static or animated worked examples? *Applied Cognitive Psychology* 21, 6 (2007), 747–764.

M. K. McCluskey and K. E. Cullen. 2007. Eye, head, and body coordination during large gaze shifts in rhesus monkeys: Movement kinematics and the influence of posture. *Journal of Neurophysiology* 97, 4 (2007), 2976–2991.

A. Mehrabian. 1966. Immediacy: An indicator of attitudes in linguistic communication. *Journal of Personality* 34, 1 (1966), 26–34.

A. S. Meyer, A. M. Sleiderink, and W. J. M. Levelt. 1998. Viewing and naming objects: Eye movements during noun phrase production. *Cognition* 66, 2 (1998), B25–B33.

J. Mumm and B. Mutlu. 2011. Designing motivational agents: The role of praise, social comparison, and embodiment in computer feedback. *Computers in Human Behavior* 27, 5 (Sept. 2011), 1643–1650.

D. L. Mumme, E. W. Bushnell, J. A. DiCorcia, and L. A. Lariviere. 2007. Infants' use of gaze cues to interpret others' actions and emotional reactions. In *Gaze Following: Its Development and Significance*. Lawrence Erlbaum Associates, 143–170.

B. Mutlu, J. Forlizzi, and J. Hodgins. 2006. A storytelling robot: Modeling and evaluation of human-like gaze behavior. In *Proceedings of 6th IEEE-RAS International Conference on Humanoid Robots*. IEEE, 518–523.

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

Y. I. Nakano, G. Reinstein, T. Stocky, and J. Cassell. 2003. Towards a model of face-to-face grounding. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1 (ACL'03)*. Association for Computational Linguistics, 553–561.

J. P. Otteson and C. R. Otteson. 1980. Effect of teacher's gaze on children's story recall. *Perceptual and Motor Skills* 50, 1 (1980), 35–42.

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

T. Pejsa and I. S. Pandzic. 2010. State of the art in example-based motion synthesis for virtual characters in interactive applications. *Computer Graphics Forum* 29, 1 (2010), 202–226.

C. Pelachaud and M. Bilvi. 2003. Modelling gaze behavior for conversational agents. In *Proceedings of 4th International Workshop on Intelligent Virtual Agents (IVA'03)*. Springer, 93–100.

J. Pelz, M. Hayhoe, and R. Loeber. 2001. The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research* 139, 3 (2001), 266–277.

D. I. Perrett and N. J. Emery. 1994. Understanding the intentions of others from visual signals: Neurophysiological evidence. *Cahiers de Psychologie Cognitive / Current Psychology of Cognition* 13, 5 (1994), 683–694.

C. Peters and A. Qureshi. 2010. A head movement propensity model for animating gaze shifts and blinks of virtual characters. *Computers & Graphics* 34, 6 (2010), 677–687.

I. Pomianowska, F. Germeys, K. Verfaillie, and F. N. Newell. 2011. The role of social cues in the deployment of spatial attention: Head-body relationships automatically activate directional spatial codes in a Simon task. *Frontiers in Integrative Neuroscience* 6, 4 (2011).

M. Scaife and J. S. Bruner. 1975. The capacity for joint visual attention in the infant. *Nature* 253, 1 (1975), 265–266.

E. A. Schegloff. 1998. Body torque. *Social Research* 65, 3 (1998), 535–596.

J. V. Sherwood. 1987. Facilitative effects of gaze upon learning. *Perceptual and Motor Skills* 64, 3 (1987), 1275–1278.

A. Tartaro. 2007. Authorable virtual peers for children with autism. In *CHI'07 Extended Abstracts on Human Factors in Computing Systems (CHI EA'07)*. ACM, New York, NY, 1677–1680.

M. Thiebaux, B. Lance, and S. Marsella. 2009. Real-time expressive gaze animation for virtual humans. In *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems - Volume 1 (AAMAS'09)*. International Foundation for Autonomous Agents and Multiagent Systems, 321–328.

T. Uemura, Y. Arai, and C. Shimazaki. 1980. Eye-head coordination during lateral gaze in normal subjects. *Acta Oto-Laryngologica* 90, 3–4 (1980), 191–198.

A. Woodward. 2005. Infants' understanding of the actions involved in joint attention. In *Joint Attention: Communication and Other Minds*. Oxford University Press, 110–128.

W. H. Zangemeister and L. Stark. 1982. Types of gaze movement: Variable interactions of eye and head movements. *Experimental Neurology* 77, 3 (1982), 563–577.

Received November 2013; revised December 2014; accepted January 2015