A Head-Eye Coordination Model for Animating Gaze Shifts of Virtual Characters

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

ABSTRACT
We present a parametric, computational model of head-eye coordination that can be used in the animation of directed gaze shifts for virtual characters. The model is based on research in human physiology. It incorporates control parameters that allow for adapting gaze shifts to the characteristics of the environment and the gaze targets and idiosyncratic behavioral attributes of the virtual character. A user study confirms that the model communicates targets as effectively as real humans, while being preferred subjectively to prior state-of-the-art models.

Categories and Subject Descriptors
H.1.2 [Models and Principles]: User/Machine Systems—Human factors; H.5.2 [Information Interfaces and Presentation]: User Interfaces—Evaluation/methodology, User-centered design

General Terms
Algorithms, Design, Experimentation, Human Factors

Keywords
Eye-gaze behavior, embodied conversational agents, empirical study

1. INTRODUCTION
Gaze shifts—the intentional redirection of gaze toward a particular piece of information in the context of interaction—are a fundamental building block of overall human gaze behavior. Through subtle variation in timing and movement of the head and eyes in pointing the gaze, individuals construct a range of complex communicative behaviors. When animating a virtual agent, control mechanisms must synthesize the wider range of such movements so that the agent displays natural communicative behaviors, yet provide sufficient control over the subtleties of the movements to allow for individual variation and expressions. Creating control mechanisms that achieve the combination of communicative effectiveness, naturalness, and parametric control remains an open challenge.

In this paper, we present a parametric control model that can be used to animate gaze shifts for virtual characters. Our model builds on findings from neurophysiology and procedurally specifies combined head and eye movements to create humanlike gaze shifts. The physiological basis of our model helps achieve effective communication and subjective naturalness, while the procedural implementation allows for parametric control over gaze shifts generated by the model. An empirical study confirms that the model meets our goals of creating gaze shifts that effectively communicate gaze direction and appear natural and realistic. In addition, recent work [1] has shown that the parametric control this model provides makes it capable of achieving high-level outcomes such as affiliation and learning.

2. BACKGROUND
This section reviews existing models for gaze synthesis from computer graphics and human-computer interaction (HCI), as well as the neurophysiological research that informs our gaze model.

2.1 Computer Graphics and HCI
Numerous gaze models have been proposed in the literature, each with different methods, goals, and contributions. For example, data-driven models [3, 11] are a common and powerful approach. However, it is often difficult to manipulate parameters or incorporate known constraints in these models without providing new hard-to-find or hard-to-create examples. The “expressive gaze model” [16] takes a hybrid data-driven/procedural approach, focusing on the communication of emotion. Because this approach handles the head and eyes separately, it does not cover the complexities of eye-head synchronization during gaze shifts as our model does. Other models consider where the agent should be looking [12, 13], explain why an agent might be looking toward particular targets [6, 17, 22], or generate idle gaze behavior when the agent is not actively gazing [2, 18].

A current state-of-the-art model takes a similar approach to ours, procedurally generating gaze shifts while taking idiosyncratic head propensity into account [21]. Head propensity is defined in this model as the amount the head is employed in generating gaze shifts as our model does. Other models consider where the agent should be looking [12, 13], explain why an agent might be looking toward particular targets [6, 17, 22], or generate idle gaze behavior when the agent is not actively gazing [2, 18].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

4th Workshop on Eye Gaze in Intelligent Human Machine Interaction: Eye Gaze and Multimodality Santa Monica, California, USA
Copyright 2012 ACM X-XXXX-XX-X/XX/XX ...$15.00.
tity model—does not involve head latency and has not been shown to accommodate the wider range of gaze behaviors. Our evaluation confirms that our model generates more natural and realistic gaze shifts than the head propensity model. Furthermore, our model has been proven to be capable of achieving high-level outcomes such as improved affiliation with the agent and improved learning through principled manipulations of the subtle low-level variables provided by the model [1].

2.2 Neurophysiology

Research in neurophysiology has studied how humans carry out gaze shifts by coordinating head and eye movements in a tightly connected dynamic process. In most existing models of directed gaze shifts, kinematics of saccadic movements, such as duration and peak velocity, are simplified, producing movements that depend only on the amplitude and direction of directed eye movements towards the target. In reality, concurrent head movements affect eye movements significantly [5]. This causal relationship holds in both directions; for example, head movement amplitude decreases as the eyes 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 the eyes to increase their contribution to the shift [19].

Head movement latency in relation to the onset of the eye movement can vary from person to person and task to task. Factors such as target amplitude, the predictability of the target, target saliency, vigilance of the subject, and whether the gaze shift is forced or natural affect this latency [20, 23]. The modality of the gaze target makes a difference as well; studies that compare auditory and visual targets show that eyes tend to lead the head most often when people orient to visual targets, whereas the head tends to lead the eyes most often when people orient to auditory targets [8, 9].

Humans are mechanically limited in their ability to rotate their eyes, a limitation referred to as the oculomotor range (OMR). The human OMR has been estimated to be between 45° and 55°. However, merely encoding these OMR values as static parameters into virtual humans is not sufficient, as the effective OMR may fluctuate during the course of a single gaze shift. The fluctuation is a product of the neural (as opposed to mechanical) nature of the limitation imposed on eye motion [10].

The degree to which individuals use their heads in performing a gaze shift is highly idiosyncratic. The physiological research literature describe some people as “head-movers,” i.e., individuals who move their head fully to align with the gaze target every time, and some as “non-movers” [7]. From a bio-mechanical standpoint, humans should universally be “non-movers,” as fully moving the head—which is almost a hundred times heavier than the eyes—is not an economical solution [14]. This idiosyncratic characteristic of head and eye movements has a neural basis, and is often not captured by a naive inverse kinematics solutions for animating gaze shifts.

3. MODEL OF HEAD-EYE COORDINATION

In the model we present here, the parameters of the specific gaze shift (e.g., the target direction) and parameters of the character (e.g., maximum head velocity) are used to compute a number of internal timing parameters at the onset of the gaze movement. Table 1 lists all parameters. Once the internal parameters are computed, the gaze shift begins, and eyes and the head are rotated directly towards the target at dynamically-changing angular velocities. The velocities are recomputed in each frame of animation, based on gaze shift progress and current rotations of the eyes and the head. This allows the model to react to perturbations of the head position or target during motion. The model calculates the rotations for the eyes independently in order to achieve convergence. A visual representation of this model is provided in Figure 1.

3.1 Internal Parameter Computation

The first phase in generating gaze shifts is to determine the latency of the onset of head movement, $hl$, in relation to the onset of the eye movement (Figure 1b). Whether an individual follows a head-first or an eyes-first approach is determined by factors such as the vigilance of the agent ($AG\text{vig}$), the target salience ($GT\text{sal}$), the eccentricity of the target ($G\text{Amp}$), the predictability of the target location ($GT\text{pred}$), and the intent of the agent—forced or natural shift ($AG\text{int}$) (Figure 1a). Each of these factors is associated with a different likelihood ratio of leading to a head-first versus eyes-first gaze shift. These ratios are summarized in Table 2. For example, gaze shifts with large amplitudes (greater than 30°) are 3.05 times more likely to involve a head-first approach. When considered in isolation from the other parameters, auditory targets always produce head-first gaze shifts, hence the ratio value of $\infty$ [23].

A ratio value $r$ is defined as $\frac{P_h}{P_e}$, where $P_h$ is the probability of a head-first gaze shift. Rewriting for $P_h$, we get

$$P_h = \frac{r}{r+1}.$$

We can compute the final probability of a head-first gaze shift by linearly interpolating between the probability of a head-first gaze shift ($P_h$) and an eyes-first gaze shift ($1 - P_h$). Probabilities for each factor are sampled to determine whether each suggests a head-first or an eyes-first movement. These various “votes” on the type of movement are combined to determine the head latency. If $f$ is the number of factors that vote to determine a head-first gaze shift and $n$ is the total number of parameters, then the ratio $s$ can be computed as $\frac{r}{s}$. We can use $s$ to compute the head latency ($hl$) for the gaze shift by linearly interpolating between the head latency of a purely head-first gaze shift, $hls$, and the head latency of a purely eyes-first gaze shift, $hle$. We chose to use 100 ms for $hls$ and 100 ms for $hle$, in the implementation of our model based on the range of values proposed in the neurophysiology literature.

3.2 Generating Gaze Motion

Once the model determines the head latency, it initiates the movements of the eyes and/or the head. Each eye and the head move towards the target, following velocity profiles that resemble standard ease-in and ease-out functions (Figure 1c). These movements prevent unnatural head motions caused by high-frequency signals. The maximum velocity of the eyes and head, $V_{max}$, are computed based on positive linear relationships with the amplitude of the intended gaze shift [10]. We derived a piecewise polynomial function to approximate the full velocity profile for both the head and the eyes determined from the literature [18, 14]. This
polynomial function can be expressed as follows, where $g$ is the proportion of the gaze shift completed, $V_{\text{max}}$ is the maximum velocity, and $V$ is the current calculated velocity.

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

Human eye rotations have limitations defined by the oculomotor range (OMR) (Figure 1d). A virtual character’s baseline OMR can be empirically determined based on the size of the eye cavities of the character model. At the onset of a gaze shift, $OMR_{\text{eff}}$—a neurally determined limit on eye motion—is computed based on the initial eye position ($IEP$) and the OMR. $IEP$ is measured in degrees as a rotational offset of the current eye orientation from a central (in-head) orientation. This value is only non-zero when the rotational offset is contralateral (on the opposite side of center) to the target. When the eyes begin the gaze shift at these angles, the $OMR_{\text{eff}}$ has a value close to the original baseline value. When the eyes begin the gaze shift closer to a central orientation in the head, the $OMR_{\text{eff}}$ diminishes [4]. We approximated this relationship in our implementation with the following function:

$$OMR_{\text{eff}} = OMR \cdot \left( \frac{1}{360} IEP + 0.75 \right).$$

$OMR_{\text{eff}}$ is also updated throughout the gaze shift at every time step according to the concurrent head velocity, $HV$. As the head moves faster, the $OMR_{\text{eff}}$ diminishes [10]. This relationship was approximated in our implementation with the following function, where $OMR_{IEP}$ is the value for $OMR_{\text{eff}}$ that was computed in the previous equation at the onset of the gaze shift.

$$OMR_{\text{eff}} = OMR_{IEP} \cdot \left( \frac{1}{600} HV + 1 \right)$$

Head alignment is a user-defined parameter which specifies a significant amount of idiosyncratic variability in the behaviors of the character, namely, whether the character is a “head-mover” or a “non-mover” (Figure 1e). A parameter value of 0% for head alignment indicates that once the eyes have reached the gaze target, the head stops moving. On the other hand, at a 100% parameter value for head alignment, the head continues rotating until it is completely aligned with the target. Head alignment values between these two extremes can be computed using a linear interpolation between the two corresponding rotational values.

---

### Table 1: All parameters of the model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type</th>
<th>Symbol Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaze Target Position</td>
<td>Input</td>
<td>$GT_{\text{pos}}$</td>
</tr>
<tr>
<td>Gaze Amplitude</td>
<td>Internal</td>
<td>$GA_{\text{mp}}$</td>
</tr>
<tr>
<td>Target Predictability</td>
<td>Input</td>
<td>$GT_{\text{pred}}$</td>
</tr>
<tr>
<td>Target Saliency</td>
<td>Input</td>
<td>$GT_{\text{sat}}$</td>
</tr>
<tr>
<td>Target Modality</td>
<td>Input</td>
<td>$GT_{\text{mod}}$</td>
</tr>
<tr>
<td>Agent Vigilance</td>
<td>Input</td>
<td>$AG_{\text{vig}}$</td>
</tr>
<tr>
<td>Agent Intent</td>
<td>Input</td>
<td>$AG_{\text{int}}$</td>
</tr>
<tr>
<td>Head Latency</td>
<td>Internal</td>
<td>$hl$</td>
</tr>
<tr>
<td>Oculomotor Range</td>
<td>Input</td>
<td>$OMR$</td>
</tr>
<tr>
<td>Effective OMR</td>
<td>Internal</td>
<td>$OMR_{\text{eff}}$</td>
</tr>
<tr>
<td>Head Alignment</td>
<td>Input</td>
<td>$h_{\text{align}}$</td>
</tr>
<tr>
<td>Maximum Eye Velocity</td>
<td>Internal</td>
<td>$EV_{\text{max}}$</td>
</tr>
<tr>
<td>Eye Velocity</td>
<td>Internal</td>
<td>$EV$</td>
</tr>
<tr>
<td>Maximum Head Velocity</td>
<td>Internal</td>
<td>$HV_{\text{max}}$</td>
</tr>
<tr>
<td>Head Velocity</td>
<td>Internal</td>
<td>$HV$</td>
</tr>
<tr>
<td>Initial Eye Position</td>
<td>Input</td>
<td>$IEP$</td>
</tr>
</tbody>
</table>
The last phase of the model involves the vestibulo-ocular reflex (Figure 1f). When the eyes reach the gaze target, they remain locked to the target for the remainder of the gaze shift, rotating in the opposite direction of head motion as the head completes its portion of the gaze shift.

Ancillary Components of Our Model – Our model also includes a blink controller that serves two key functions: generating gaze-evoked blinking as described by Peters [21] and idle blink behavior. In addition, when the agent is not actively engaging in a gaze shift following our model, i.e., when it is in idle gaze state, the eyes are controlled by an implementation of the model presented in Lee et al. [18]. This implementation creates subtle random eye movements in a principled way and dramatically increases the realism of the character.

4. EVALUATION

Development of our model was followed by an empirical evaluation of the communicative accuracy and perceived naturalness of gaze shifts generated by it. We conducted a study to compare gaze shifts generated by our model against those generated by a state-of-the-art model, as well as against gaze shifts displayed by human confederates. In addition we explored the effect that participant gender and the gender of the virtual character had on our measures.

Experimental Setup and Task – Participants were shown a series of videos in which either an animated virtual character or a human confederate shifted gaze toward one of sixteen objects arranged on a desk. This simplified scenario allowed us to focus our evaluation on the effectiveness and naturalness of gaze shifts, while minimizing contextual and interactive 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, distinguished by color and shape. Still images from the videos are shown in Figure 2.

Study Design – We conducted a 2x2x8 factorial experiment with split-plot design. Our factors included participant gender (two levels varying between participants), gender of the virtual agent (two levels varying between participants) and model type (eight levels varying within participants). The model type independent variable included comparison conditions for gaze shifts generated by the head propensity model [21] and those produced by our model. For each model, we manipulated the model parameters to create three distinct comparison conditions with different head alignment/propensity levels, 0%, 50%, or 100%, with the goal of determining how changes in the model parameters affected the communicative accuracy and perceived naturalness of the gaze shifts.

The model type independent variable also included two control conditions. In the first control condition, a male or female human confederate presented gaze shifts toward the object on a desk in front of him/her. The second control condition involved a virtual agent maintaining gaze toward the participant without producing any gaze shifts.

Study 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 are 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 filled out a subjective evaluation. Participants were exposed to each model condition four times in a stratified order. This ordering ensured that the 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.
Communicative Accuracy

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 crowd-sourcing best practices to minimize the risk of abuse and to achieve a wide range of demographic representation [15]. Participants received $2.50.

Measurement – 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 directed. To measure perceived naturalness, we constructed a scale using five items that measured naturalness, human-likeness, lifelikeness, realism, and expressiveness. Participants rated each video for each item using a seven-point rating scale. A confirmatory factor analysis of the data collected from these items showed that the item “expressive” had a low loading on the scale (r = .45). We excluded this item from the resulting scale, producing a highly correlated four-item scale of perceived naturalness, Cronbach’s α = .94.

4.1 Results

We conducted a mixed-model analysis of variance (ANOVA) on the data from the first study to determine the effect that different gaze models had on how accurately participants identified the object that the agents or the human confederates looked toward and the perceived naturalness of the gaze shifts. Overall, model type had a significant effect on communicative accuracy, $F(7, 46.67) = 16.77, p < .001$, and perceived naturalness, $F(7, 46.67) = 151.03, p < .001$. Detailed comparisons across conditions for each factor are described in the next paragraphs.

Communicative Accuracy – Pairwise 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, $F(1, 14.91) = 0.03, p = n.s$. Similarly, no differences in accuracy were found between our model and the head propensity model, aggregated across all levels of head propensity, $F(1, 1967) = 3.34, p = .068$. Comparisons over the realism scale (one of the items included in the perceived naturalness scale) found that gaze shifts produced by our model were rated as significantly more realistic than those generated by the head propensity model, $F(1, 1963) = 5.75, p = .017$. Finally, 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, $F(1, 486.1) = 9.40, p = .002$. Results on the communicative accuracy and perceived naturalness measures are illustrated in Figure 3.

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, $F(1, 851.6) = 17.17, p < .001$. 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, $F(1, 847.1) = 4.85, p = .028$. Finally, the analysis found a marginal interaction between participant gender and the gender features of the virtual character, $F(1, 843.3) = 3.20, p = .074$. Figure 3 illustrates these results.

The results from these studies show that gaze shifts generated by our model communicate gaze direction as accurately as do human gaze shifts and those generated by the state-of-the-art model. They also show that gaze shifts generated by our model are perceived as significantly more realistic and marginally more natural than those generated by the state-of-the-art model. In addition, recent work has shown that the parametric control this model provides makes it capable of achieving significant outcomes, such as affiliation and learning [1]. Finally, our results indicate that character design, particularly gender-based features, might have an effect on the perception of the gaze shifts performed by virtual characters.

![Figure 2: Still images from the videos presented to the participants.](image)

![Figure 3: Results from the communicative accuracy, perceived naturalness, and realism measures. The baseline model refers to the head propensity model we used for comparison.](image)
REFERENCES


