A Model of Gaze for the Purpose of Emotional Expression in Virtual Embodied Agents

Brent J. Lance
University of Southern California
Information Sciences Institute
4676 Admiralty Way Suite 1001
Marina Del Rey, CA 90292
lance@usc.edu

Stacy C. Marsella
University of Southern California
Information Sciences Institute
4676 Admiralty Way Suite 1001
Marina Del Rey, CA 90292
marsella@isi.edu

ABSTRACT
Currently, state of the art virtual agents lack the ability to display emotion as seen in actual humans, or even in hand-animated characters. One reason for the emotional inexpressiveness of virtual agents is the lack of emotionally expressive gaze manner. For virtual agents to express emotion that observers can empathize with, they need to generate gaze - including eye, head, and torso movement - to arbitrary targets, while displaying arbitrary emotional states. Our previous work [18] describes the Gaze Warping Transformation, a method of generating emotionally expressive head and torso movement during gaze shifts that is derived from human movement data. Through an evaluation, it was shown that applying different transformations to the same gaze shift could modify the affective state perceived when the transformed gaze shift was viewed by a human observer. In this paper we propose a model of realistic, emotionally expressive gaze that builds upon the Gaze Warping Transformation by improving the transformation implementation, and by adding a model of eye movement drawn from the visual neuroscience literature. We describe how to generate a gaze to an arbitrary target, while displaying an arbitrary emotional behavior. Finally, we propose an evaluation to determine what emotions human observers will attribute to the generated gaze shifts. Once this work is completed, virtual agents will have access to a new channel for emotionally expressive behavior.

Categories and Subject Descriptors
I.3.7 [**Computer Graphics**]: Three Dimensional Graphics and Realism – *Animation*.

General Terms
Human Factors.

Keywords
Gaze, Nonverbal, Emotional Expression, Animation, Motion Capture, Posture, Virtual Agent

Cite as: A Model of Gaze for the Purpose of Emotional Expression in Virtual Embodied Agents, Lance, B., Marsella, S., *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Padgham, Parkes, Müller and Parsons (eds.), May, 12-16., 2008, Estoril, Portugal, pp. XXX-XXX.

1. INTRODUCTION
Currently, state of the art virtual embodied agents are deficient in emotional expressivity when compared to animated characters in feature films. Characters in animated feature films can come alive, enthrall audiences and critics, and even win Oscar awards. Virtual embodied agents currently cannot. Attempts to address this deficiency in expression have led researchers to add capabilities such as facial expression [28], increasingly realistic voices [8], prosody, head movement [5], and gesture [15] to virtual agents. Despite these additional capabilities, the expressivity gap remains. One notable aspect of film characters is the ability of the animator to reveal important information about the character's emotional state through the use of mere glances or glares while the character remains silent. This ability suggests that one of the reasons for the continued gap between virtual agents and animated characters, despite the increasing similarity in capability, is the lack of a model of emotionally expressive gaze manner for virtual characters.

However, providing a model of emotionally expressive gaze manner is a significant challenge. In order to provide a virtual agent with the same gaze capabilities as a human or an animated character, the agent must be able to perform a gaze shift to an arbitrary target, while displaying an arbitrary emotional state. But, the field of psychology has yet to fully describe how humans express emotion through their eyes. In addition, gaze consists of more than just eye movement. A gaze shift can also include movement of the head and torso, all of which must be integrated so that the overall movement is natural, and the emotional display is coherent.

An added difficulty is that an improper implementation of a model of emotionally expressive gaze may be worse than none at all. A lack of appropriate physical interrelations between distinct body components leads to robotic and unnatural movement, and the behaviors then generated seem random and disjointed. For example, the film "The Polar Express," had motion-captured characters, but the eyes were animated separately using traditional animation techniques. Several reviewers pointed out the unnatural relationship between the eyes and the body, referring to the animated characters as "zombies." One reviewer went so far as to explicitly state "Although the human characters look about 90% lifelike, it is that darn 10% (mostly the lifeless eyes) that winds up making them seem really creepy," [30].

In this research, we are interested in discovering how to create a virtual human that is capable of using the manner of its gaze behavior to express its emotional state. We are building upon existing work on Gaze Warping Transformations (GWTs), which are parameter sets for generating emotionally expressive head and torso movement during gaze shifts. GWTs are derived from the difference between human motion data of emotionally neutral and emotionally expressive gaze shifts. Once generated, they can be applied to new emotionally neutral gaze shifts, to generate new emotionally expressive gazes.

We will describe several improvements to the GWT that increase the quality of the animations produced using it. Next, we will cover how we are modeling eye movement, and how we integrate the eye movement with the Gaze Warping Transformations, using stereotypical eye movements based on the visual neuroscience literature. Finally, we describe a proposed model of expressive gaze behavior along with the intended evaluation for that model. We have based our model of expressive gaze manner firmly in human behavior by generating our movement with motion capture data, and drawing both our model of eye movement, and our integration of eye movement and the motion capture from work in the visual neuroscience literature, such as [20]. Therefore, we will be able to avoid the problems exemplified by the movie "The Polar Express."

2. RELATED WORK

Gaze is a vital part of the human sensory apparatus. As such, there has been considerable research into the mechanical properties and neural control of not only the eyes, but how the eyes and the head interact to perform gaze [12], [13], [20], [31], [35]. This research has led to a fairly complete picture of the way in which normal eye and gaze movements are performed. However, in addition to being a vital part of the human sensory apparatus, gaze is important to interpersonal interaction.

Even though there are many communicative signals that can be sent through the manner of an individual's gaze behavior [2], [16], our interest is in using gaze to express a virtual agent's emotional state. In our previous work on emotionally expressive gaze, [17], we found a greater association between emotional states and a set of physical parameters describing gaze movement, than we did for other communicative signals. We also found that the manner, the way in which gaze was performed was a highly expressive signal. Further research demonstrated that modifying the physical parameters describing the way in which a virtual agent's head and body moved during gaze shifts would change the emotional state attributed to the virtual agent by human observers [18].

Some of the affective states we are specifically interested in are dominance / submissiveness, arousal / relaxation, and pleasure / displeasure. For example, dominance can be signified through gaze behavior [9], head movement [25], or even posture [6]. A steady gaze and upright posture signal dominance, while the opposite can signal submission [6]. Arousal and gaze are also closely related [2], further demonstrated by work showing that arousal could be detected through the velocity of a gaze shift [18]. However, unlike dominance and arousal, pleasure may not be displayed through gaze at all [16], though it is through the posture of the head [25] and the body [7]. In addition to the display of these more abstract affective states, there has also been research on how head and body posture display specific emotional categories, such as in [7], [24], [33].

There have also been many implementations of gaze behavior in virtual agents, although few of these have touched on using gaze to display emotion, such as [10], who used an animated pair of eyes to display affective signals to observers. Most gaze models instead focus on sending communicative signals through gaze [3], [27], task performance and environmental interaction [29], or resting gaze, when the eye performs no other tasks [19]; although [14] used a procedural model of eye movement based in neuroscience to animate a virtual agent that regulated gaze direction through direction attention at visually salient points.

In addition to research on gaze behavior in humans and virtual agents, our work is inspired by research on transferring parameters that describe the "manner" or "style" in which a movement is performed from one movement to another. The primary work on style uses "style machines," which are combinations of Hidden Markov Models, to learn a style, such as "anger" from one movement and apply it to another [4]. Other research has dealt with modifying the style of walking animations [21], using movement manner to display emotion in hand gestures through learning Laban movement parameters with neural networks [37], or by modifying the gesture with emotional transforms [1]. The model of a Gaze Warping Transformation is similar to [1]'s Emotional Transforms for gesture. However, the Gaze Warping Transformation provides a greater resolution for applying the transform to the original movement and we explored how the transform, once applied, changes the perception of an observer [18].

While there has been much research on eye movement, gaze behavior, emotional expression, and the transfer of manner, there has to date been no exploration of specifically how changes in emotional state affect changes in the manner of gaze behavior. This paper describes a model of gaze behavior intended to be used as a basis for performing that exploration.

3. MODEL OF EMOTION

As a framework for our implementation of emotionally expressive gaze, we are using the PAD model of emotion [22]. The PAD model is a three dimensional model of emotion, and the acronym "PAD," comes from the Pleasure, Arousal, and Dominance dimensions that make up the model. The more commonly-known emotions such as anger and fear are represented in the PAD model as sub regions of the space defined by the three dimensions. For example, anger is represented as low pleasure,

Table 1: Emotional Categories and PAD Dimensions

Emotional Category	PAD Dimension Associations
Anger	-P +A +D
Disgust	-P-A+D
Fear	-P+A-D
Sadness	+P+A+D
Shame	-P-A-D
Happiness	-P+A-D
Admiration	+P+A-D

high arousal, and high dominance, while fear is represented as low pleasure, high arousal, and low dominance.

We are using the PAD model for several reasons. One reason is that there is a body of research describing how gaze behavior varies along each dimension of the model [2], [9], [16]. We have also had some success with an implementation of gaze that expressed the Arousal and Dominance dimensions [18]. Another reason for using the PAD model is that the dimensions of the model were derived in part from factor analyses of nonverbal behavior studies [23]. This indicates that it may be possible to generate gaze behavior expressing specific emotional states by combining behaviors associated with the PAD dimensions. The categorical model of emotion that we are using to map into the PAD space is shown in Table 1. These categories are loosely based on the mapping of OCC emotional categories to PAD space as described in [11]. If gaze behavior can be generated by using the PAD dimensions to determine behavior for emotional states, this makes the process of generating gaze behavior that expresses arbitrary emotional states simpler. For this purpose, we intend to explore the mapping between PAD dimensions and physical behaviors, as described in the evaluation section below.

4. MODEL OF EXPRESSIVE GAZE

4.1 Previous Work: the GWT

Our previous work describes the Gaze Warping Transformation (GWT), a method for generating emotionally expressive head and torso movement during gaze shifts [18]. The GWT is a combination of temporal scaling and spatial transformation parameters (as described in [34]), that is used to transform an emotionally neutral gaze shift towards a target into an emotionally expressive gaze shift directed at the same target. A small number of transformations would then be used to produce various gazes covering a variety of emotional expressions directed towards arbitrary targets

4.1.1 Movement Representation

Animated movements are usually represented as a set of motion curves, where each curve represents the value of a single degree of freedom in the animated body over the number of frames in the animation. A common sparse movement representation is one using "keyframes." The keyframes of an animation are a subset of the frames for that animation, such that the values of the animation curves for the intermediate frames can be found by interpolating between keyframe values.

4.1.2 Finding the Gaze Warping Transformation

A GWT is found by obtaining two motion captures of gaze shifts directed from the same start point to the same target, one emotionally expressive, the other emotionally neutral, and finding the Motion Warping [34] parameters that would convert the animation curve representing each degree of freedom in the emotionally neutral animation into the animation curve for the corresponding degree of freedom in the emotionally expressive movement, similar to [1].

A Motion Warping curve provides a transformation from the keyframes of an animation curve $x(t)$ defined as a set of value, frame pairs, (x_i, t_i) to those of a new motion $x'(t')$ defined as the set of pairs (x'_i, t'_i) through the use of two functions [34]. The first function, given t'_i, calculates the frame t_i in the neutral motion curve to obtain x_i. We use the function

$$t_i = g(t'_i)$$

$$g(t'_i) = c(t'_i) * t'_i$$

where given an actual frame time t'_i, in the emotional movement, $g(t)$ determines the corresponding frame t_i in the neutral movement through a time scaling parameter $c(t'_i)$. The other function is

$$x'(t_i) = x(t_i) + b(t_i)$$

where $b(t_i)$ is a spatial offset parameter that transforms $x(t_i)$ into $x'(t_i)$. The final GWT is an $m * n$ set of (c, b) pairs, where m is the number of degrees of freedom in the animated body, and n is the number of keyframes in the animation.

4.1.3 Applying the Gaze Warping Transformation

Applying the GWT is done by using the parameters that make up the GWT to modify the keyframes of an emotionally neutral gaze shift into one which displays the same emotion as the original shift. The first step is to downsample the neutral animation to have the same number of keyframes as the GWT. Then, the GWT is applied to time scale and spatially offset the keyframes of the neutral animation. Finally, the intermediate frames are recovered using cubic interpolating splines, generating a new emotionally expressive gaze shift.

The GWT is evaluated by displaying animations of the generated gaze shifts to human observers. The animated body that displays the gaze shifts is very simple, lacking facial expression or even arms, in order to ablate nonverbal behavior not related to the gaze shift. The observers rate the animations based on the emotion expressed. In [18], observers rated the animations in two separate ways: first, they compared two animations, and selected the one which more closely fit a specific description. Second, they rated each animation individually on a Likert scale by how well it fit the same description. By analyzing the results of these ratings, it can be determined if applying different GWTs to the same emotionally neutral animation results in observers attributing a different emotional state to the resulting animations.

The motion captures used as a basis for the GWT were obtained using three Ascension Flock of Birds sensors: two sensors capture the position and orientation of the torso, and a third sensor captures the position and orientation of the head. In addition, an ASL head mounted eye tracking rig recorded the movement of the eyes, although this data was not integrated into the GWT.

4.2 Current Work: Improving the GWT

As the GWT is based on a technique of simple geometric transformations the animations generated using it would often contain artifacts. Two types of artifacts were seen in the animations. First, discontinuities appeared in the movement, where the animated character would "pop" to a new position without moving through the space to that position. Second, body parts moved outside the physical limits of an actual human body.

4.2.1 GWT Improvements

In order to solve these problems, we have improved several features of the GWT. The GWT is applied at keyframes of the original movement curve, and the intermediate points are obtained through interpolation. However, the least-squares-difference algorithm that was used to obtain the keyframes could result in different sets of keyframes being obtained from different movement curves. The effect of this is that a transformation derived from a specific location in one gaze shift would be applied to a nearby, but not corresponding location on another gaze shift. These misapplications led to movement discontinuities.

In order to ensure that the keyrframes the GWT is applied to match across different curves, we align all of our gaze shifts to a "stereotypical gaze shift," using an alignment algorithm derived from that described in [1]. This algorithm aligns the curves of two animations or motion captures based on the ratio of movement which has occurred by a specific time to that which occurs throughout the entire curve. The values used for alignment are found by the function

$$\hat{f}(t) = \frac{\sum_{\tau=0}^{t} |\mathbf{v}(\tau)|}{\sum_{\tau=0}^{t_{end}} |\mathbf{v}(\tau)|}$$

where **v(t)** is the three dimensional velocity vector of the head. Thus, the movement of the head, which is generally of higher amplitude than that of the torso during a gaze shift, is used as an approximation for movement of the entire body. For example, if we wish to align two movement curves, the first of which has 20% of its head movement occur by frame 50, while the other has 20% of its head movement occur by frame 75, then we can align those two frames. By repeating this process for every frame in both curves, we can align the curves much more accurately. We then select the keyframes based on this alignment with the "stereotypical" gaze shift, ensuring that transformations are then applied to corresponding locations.

In order to remove the problem of body parts moving outside the limit of a human body, we added an inverse kinematics system implemented using nonlinear optimization, similar to [36]. This system simulates a rigid skeleton in place of the independent data received from each motion tracking sensor. This provides improved performance when compared to an ad-hoc set of constraints, and keeps the various components of our animated body within the limits of the human body. These two additions have vastly improved the quality of the animations obtained through using the GWT.

4.2.2 Generating Neutral Head and Body Movement

In order to generate emotionally expressive gaze shifts containing head or torso movement to an arbitrary target, we use GWTs to transform emotionally neutral gaze shifts directed towards that target into emotionally expressive gaze shifts. Therefore, we must first obtain the neutral gaze shift. We are currently using GWTs to generate these neutral gaze shifts as well. For this purpose, we have collected a set of angular GWTs that warp a stationary body by rotating it to perform a gaze shift at a known target. In order to perform a gaze shift from an arbitrary location to an arbitrary location, we find the angular difference between the start and end positions, and apply the appropriate angular GWT to the stationary position held by the character. We use this procedure because it maintains coherency with our model of emotionally expressive gaze, although the emotionally neutral gaze shifts to arbitrary targets could be generated in any fashion.

4.3 Model of Eye Movement

4.3.1 Types of Eye Movement

In addition to the improved GWT, which describes movement of the head and body during gaze shifts, we have also developed an integrated model of eye movement, based in part on the visual neuroscience literature. Originally, our intent was to use a head-mounted eye tracker to collect eye data, and base our model of eye movement off of that collected data. However, the head-mounted eye tracker lost track of the eye position in many common situations. For example, it lost track when the eye moved more than 15º away from looking straight ahead, or if the head moved too quickly. Further, we found that the data we did collect matched very well with the stereotypical eye movements described in the visual neuroscience literature, which used more intrusive, but much more accurate eye tracking technology. Therefore, we have chosen to implement our model of eye movement using these stereotypical movements, instead of directly using eye movement data.

In [20]'s overview of visual neuroscience, the authors identify several different functional classes of eye movement. Those which are relevant to our work are the vestibular and optokinetic, both of which maintain visual fixation during head movement; and saccades, which are rapid and highly stereotyped eye movements to a specific target. In addition, they discuss the combination of head and eye movement, through the existence of eye-head saccades. Thus, our model implements vestibulo-ocular reflex (VOR) movement, saccades, and combined eye-head saccades. In addition, we differentiate VOR with head movement from VOR with head and torso movement, and we similarly differentiate between combined eye-head saccades and eye-head-torso saccades. Currently we are unsure if this distinction between gaze shifts using just the head, and gaze shifts performed using the head and the torso is necessary, but we will be testing it as a part of our planned evaluation.

4.3.2 Eye Movement Representations - Saccade

Figures 1 through 4 show how eye movements are represented for each of the different types of gaze shift. In each figure, the Gaze Angle curve represents the target of the entire gaze, combining eye and head orientation. The Eye Angle curve represents the angle of the eyes within their orbit, and the Head Angle curve represents the angle of the head relative to its initial position.

Figure 1 shows the motion curves for a stereotypical saccade in one dimension, and an animated figure showing the beginning and ending of the saccade. The saccade is a very rapid, highly-stereotyped eye movement which rotates the eye from its initial position directly to the target. The size, speed, and duration of saccadic movements are closely related. As the amplitude of a saccade increases, so does its velocity and duration. These relationships are called the main sequence relationships, and are

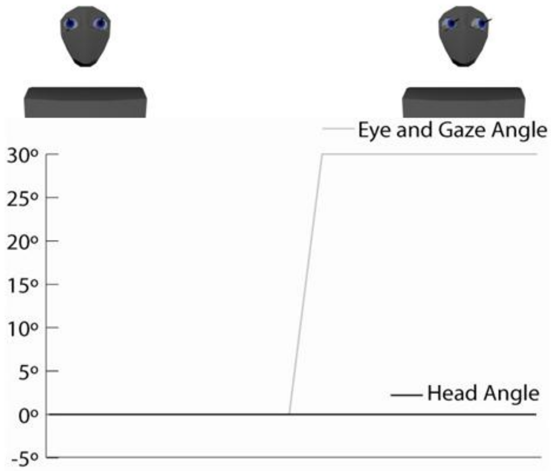

Figure 1: Stereotypical Saccade

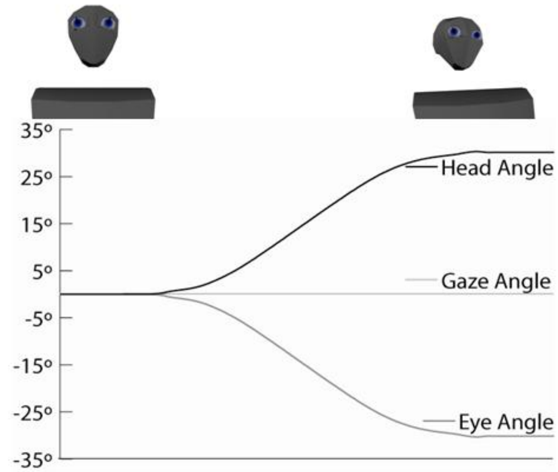

Figure 2: Vestibulo-Ocular Reflex (VOR)

used to define ranges for standard saccades [31]. Eye movement outside these ranges is either non-saccadic behavior, or an abnormal saccade, often symptomatic of pathology [20], [35].

Our representation of saccadic movement is that of a rotation to the desired target, with additional considerations: we are approximating the main sequence relationship between amplitude and duration as a linear relation between the amplitude of the saccade and the number of frames of animation the saccade takes to execute. For each ten degrees of amplitude, one intermediate frame is added. Then, linear interpolation between the start and end positions is used to determine intermediate orientation of the eye across these animation frames. The velocity is thus not directly controlled, but implicitly determined by the amplitude and duration. We have also limited the amplitude of the saccades to +-45º. While the mechanical limits of human eye movement are closer to +-55º, there is evidence showing that there are neural limits that keep eye movement within +-45º [13].

In our model, the eye saccade does not allow for concurrent head and body movement. Therefore, to generate a saccadic gaze shift, we first generate a stationary posture lasting for the desired number of frames of animation by replicating the last known posture of the animated character, with very slight perturbations to avoid the robotic appearance of utter stillness. We then use our stereotypical saccade representation to determine the location of the eyes at every frame of the animation. Finally, we use this conjoined data to drive the animated character.

4.3.3 Eye Movement Representations - VOR

Figure 2 shows the motion curve for a head movement with VOR. It also shows an animated character demonstrating the VOR movement. Through the vestibulo-ocular reflex, the eyes rotate within their orbit so that the gaze maintains the same target while the head moves. Since the VOR is a non-saccadic movement, it is not subject to the main sequence relationship, allowing the slower eye rotation that matches the head rotation. However, movement is still limited to the +-45º eye motor limit.

The VOR eye movement is implemented by first applying the desired GWT to an emotionally neutral gaze shift to generate an emotionally expressive movement. Then the eye orientation is determined by simply rotating the eyes opposite to the head rotation, so that the eyes maintain the same gaze target.

4.3.4 Eye-Head Combined Movement

The combined eye-head saccade (shown in Figure 3) is the primary way in which we integrate eye movement and head-torso movement. There are two versions of the eye-head saccade, and both are generated similarly. First, the head and body movement is generated through applying GWTs to emotionally neutral gaze shifts, and then the eye movement is generated based on the movement of the head, and the stereotypical representation of eye-head saccades. For eye-head saccades of less than 45º, the position of the eye is determined by generating a stereotypical saccade to the target once the head has turned more than 1º away from its starting location. Once the eye has reached its target, the VOR will keep the eyes on target as the head moves to the target more slowly [12], [13], [20]. The small images above the graph show the character at the beginning of the movement, then just after the saccade has occurred and the VOR is taking control of the eye movement, and then finally at the end of the gaze shift.

When looking at a target beyond the saccade limit of +-45º, the combined eye-head saccade is slightly different, as seen in Figure 4. In this case, after the emotionally expressive head and torso movement is generated, the position of the eye is determined by performing a stereotypical saccade to the maximum of 45º once the head has turned more than 1º away from its starting location. The eye will then remain in that orientation, relative to the head; until the head has rotated enough that the eye is on target. At that time, the VOR takes effect, and the eye remains on target until the head movement has terminated. The character above the diagram shows first the initial position of the gaze, then the gaze at the end of the 45º saccade, and finally the terminal position of the gaze shift. When an actual human performs an eye-head saccade, the VOR effect will often not be large enough to match the rotation of the head. In this case, very small saccades will occur, keeping the eye on target [13]. However, we have not modeled this effect, as we feel that these saccades are too small to noticeably transfer information on emotional state.

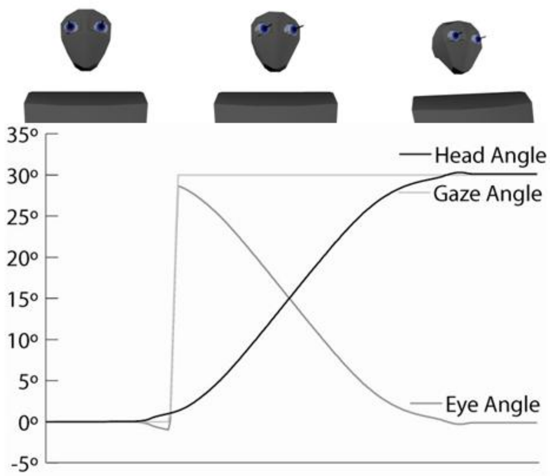

Figure 3: Eye-Head Saccade Within Eye Motor Limit

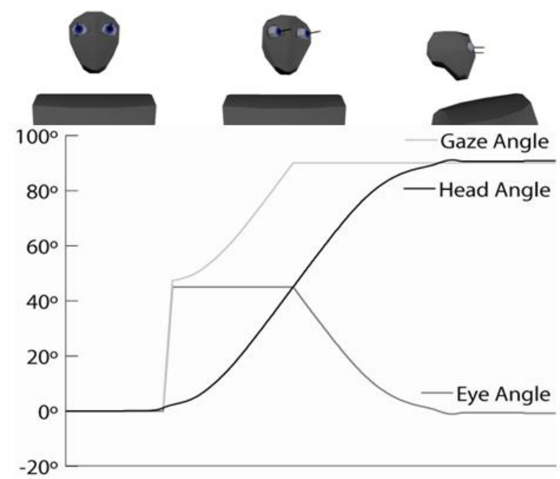

Figure 4: Eye-Head Saccade Beyond Eye Motor Limit

4.4 Expressing Emotion through Single Gazes

This model allows the generation of gaze shifts expressing an arbitrary emotion and looking at an arbitrary point in space. This is done by generating an emotionally neutral gaze shift to an arbitrary point in space, layering the emotional content on top of that, and then adding eye movement. Emotional content is layered onto head and body movement using the GWT.

However, saccades are not very expressive, as pointed out in [31]: "[a] subject's intention modifies head movement trajectories greatly, although eye saccades are quite stereotyped." In addition, informal evaluations of our model showed that modifying the speed of saccadic behavior outside that of the main sequence relationship would lead observers to attribute intoxication, not emotional state, to the resulting animations. The minor modifications of the eye movement that remain possible are unlikely to affect the impression of the character's emotional state portrayed to an observer. Therefore, one of the only remaining modifications that can be made directly to eye movement is the direction of the saccade during a gaze aversion, which has been shown to reveal some affective information [2], [10].

The limited variability of saccades means that the majority of emotional expression in a gaze shift will be due to modifications of head and torso movement by the GWT. Therefore, the main role of the eye model is to ensure realistic motion. Because the parameter space of the GWT is so large, we have identified a set of "emotional behaviors" from the psychology and arts literature that are likely to be used to reveal emotional state. This set of behaviors can be seen in Table 2.

Table 2: Combined Eye-Head Movement

Behavior	Cite	Possible Parameter Values
Head Posture	[25]	Raised, Neutral, Bowed
Torso Posture	[6, 7]	Neutral, Bowed
Movement Velocity	[26, 33]	Fast, Neutral, Slow
Attraction/ Aversion	[2, 10]	Look At, Look Away
Aversion Direction	[2, 10]	Cardinal Directions
Gaze Type	[32]	Eye Saccade, Eye-Head Saccade, VOR

In addition to identifying behaviors, we have also discretized them into a set of possible values. This was done by collecting motion captures of all of the behaviors by asking an actress to perform gaze shifts while demonstrating the requested behavior. All captured gaze shifts consist of the desired behavior being displayed first in a gaze aversion that started gazing straight ahead, and ended gazing 30° to the right, followed by a gaze attraction that turned back to straight ahead.

From these motion captures we produced GWTs representing these behaviors, by finding the difference between a gaze shift demonstrating the behavior, and a neutral gaze shift that does not. Any of these behaviors can then be displayed in a gaze shift to an arbitrary target by applying the appropriate GWT to the appropriate emotionally neutral gaze.

These behaviors were selected because they had previously been linked with a dimensional or circumplex model of emotion. For example head posture, torso posture, and direction of gaze aversion have been linked with dominance [6], [10], [25]. Velocity has been linked with arousal [26], and head posture, and gaze attractions have been linked to valence [25], [10]. However, the mapping between these behavior GWTs and either PAD dimensions or categorical emotional states is as yet unclear. This is due to a number of factors. First, we do not know exactly how these behaviors express emotion. While they have been shown to be expressive, it is not always clear what is expressed. For example, head angle can express both dominance and pleasure [25]. Second, even if we could reliably map between emotions and behaviors, it would not necessarily follow that GWTs drawn from motion captures of these behaviors would necessarily replicate those results. It is possible, for example, that our actor performs behaviors in a subtly idiosyncratic fashion, leading to a different interpretation of the behavior by an observer. Additionally, given the knowledge of how behaviors express emotion does not provide a mapping describing how combinations of those behaviors express emotion. Finally, the psychology literature does not provide the dynamics for a behavior. For example, while turning the head down can display a lack of dominance, it is unclear how much the head should be turned down, or how quickly the turn should be performed. These reasons provide the purpose for our evaluation.

Because the emphasis of this work is on generating emotional behaviors that can be correctly recognized by observers, we have focused here on the identification of a space of behaviors that can be used to express emotion in gaze, and on the ability of our model to display any of these behaviors when looking at an arbitrary point in space; not on what emotional states these behaviors actually display. We are currently working on evaluating how these behaviors, both singly, and in combination, change the perception of emotion in observers. We are particularly interested in determining how the behaviors are related to the PAD model of emotion. While we do not yet know the ideal method for combining GWTs containing different behaviors into discrete emotional states, we are also collecting motion capture data of expressions of combined behaviors, as well as discrete emotional states. We intend to use this additional data as a guide for developing our GWT combination procedure.

4.5 Expressing Emotion with Multiple Gazes

In addition to allowing the expression of emotion through individual gaze shifts, this model also allows for more subtle and complex displays of affect through sequential combinations of gaze shifts of different types, all displaying the same emotion. Consider the following comparison: first, a character that performs a single gaze shift consisting of a single eye-head saccade turning away to the side, with a slumped posture and down-turned head. In contrast, this model also allows the character to produce the same gaze shift, in the same emotional state, through the sequential combination of three individual gaze shifts. The character starts by making the same eye-head saccade shift away to the side with slumped posture and down-turned head. After turning 75% of the way to the final target, the character pauses, and shoots a quick saccade back towards the observer. At the end of this pause, the character performs a final eye-head saccade, ending at the same point as the single gaze shift described above.

5. CURRENT WORK – EVALUATION

Currently we are working on evaluating the set of emotionally expressive behaviors described above. By doing so, we intend to answer a specific set of questions. First: what emotional state do observers attribute to a character displaying each of the above behaviors? Second, what emotional state do observers attribute to a character displaying various combinations of the above behaviors? Third, can we use the PAD model to combine physical behaviors? That is, do behaviors that vary along PAD dimensions, when combined, lead to observers attributing the emotional state predicted by the PAD model to those combined behaviors? Finally, to what extent does the orientation of the gaze affect the observed emotional expression?

In order to answer these questions, we will perform an evaluation similar to that in [18]. We will be generating a large number of animated gaze shifts portraying the described emotional behaviors, and displaying these gaze shifts to human observers. By comparing the ratings by the observers of different animations, we will answer the above questions. Once we have answered these questions and integrated the information into our model of emotionally expressive gaze manner, we will be able to generate gaze shifts that display arbitrary emotional states, to arbitrary targets, by using GWTs to combine and interpolate a relatively small quantity of motion capture data.

6. CONCLUSION

In conclusion, we have described an improved version of the Gaze Warping Transformation, which provides animation of a much higher quality than earlier versions. In addition, we have provided an integrated model of eye movement, based on the physical and neurological properties of human gaze. We have also provided a model of behaviors we believe will be sufficient to express simple emotional states, and described how sequentially combining gaze shifts generated by this model can portray more subtle and complex emotional variations. Finally, we have posited a number of unanswered questions that result from this model, and described the work that we are currently performing to answer the questions and evaluate the model. We expect to find strong relationships between the various emotional dimensions and emotional states, and the various behaviors we have selected to express emotion.

7. ACKNOWLEDGMENTS

We would like to thank Dr. Bosco Tjan, Dr. Lewis Johnson, Dr. Karen Liu, Dr. Skip Rizzo, Yu-ting Ye, and Tiffany Cole for their assistance, without which this paper would not have been possible. This work was sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM), and the content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

8. REFERENCES

[1] Amaya, K., Bruderlin, A., Calvert, T. 1996. Emotion From Motion. Proceedings of the Conference on Graphical Interface, (1996). 222–229.

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

[3] Bickmore, T., Cassell, J. 2004. Social Dialogue with Embodied Conversational Agents. Bernsen, N. (ed.) Natural, Intelligent and Effective Interaction with Multimodal Dialogue Systems. Kluwer Academic (2004). 23–54.

[4] Brand, M., Hertzmann, A. 2000. Style Machines. Proceedings of SIGGRAPH, ACM Press, New York. (2000)

[5] Busso, C., Deng, Z., Neumann, U., Narayanan, S. S. 2005. Natural head motion synthesis driven by acoustic prosodic features. Computer Animation and Virtual Worlds, vol. 16, no. 3-4 (July 2005). 283–290.

[6] Carney, D. 2005. Beliefs About the Nonverbal Expression of Social Power. Journal of Nonverbal Behavior. Vol 29, No. 2 (2005). 105–123.

[7] Coulson, M. 2004. Attributing Emotion to Static Body Postures: Recognition Accuracy, Confusions, and Viewpoint Dependence. Journal of Nonverbal Behavior Vol. 28, No. 2.

[8] Douglas-Cowie, E., Campbell, A., Cowie, R., Roach, P. 2003. Emotional Speech: Towards a New Generation of Databases. Speech Communication. Vol. 40, No. 1-2. (2003). 33-60.

[9] Exline, R. 1974. Visual Interaction: The Glances of Power and Preference. Weitz, S. (ed.) Nonverbal Communication: Readings with Commentary, Oxford University Press,Oxford (1974).

[10] Fukuyama, A., Ohno, T., Mukawa, N., Sawaki, M., Hagita, N. 2002. Messages Embedded in Gaze of Interface Agents: Impression Management with Agent's Gaze. Proceedings of SIGCHI, (2002).

[11] Gebhard, P. 2005. ALMA: A Layered Model of Affect. Proceedings of AAMAS, ACM Press, New York, (2005).

[12] Gresty, M.A. 1974. Coordination of Head and Eye Movements to Fixate Continuous and Intermittent Targets. Vision Research, Vol 14. (1974). 395-403.

[13] Guitton, D., Volle, M. 1987. Gaze Control in Humans: Eye-Head Coordination During Orienting Movements to Targets Within and Beyond the Oculomotor Range. Journal of Neurophysiology. Vol 58, No. 3. (Sep. 1987). 427-459.

[14] Itti, L., Dhavale, N., Pighin, F. 2003. Realistic Avatar Eye and Head Animation using a Neurobiological Model of Visual Attention. SPIE 48th Annual International Symposium on Optical Science and Technology. 2003.

[15] Kipp, M., Neff, M., Kipp, K.H., Albrecht, I. 2007. Towards Natural Gesture Synthesis: Evaluating Gesture Units in a Data-Driven Approach to Gesture Synthesis. Proceedings of Intelligent Virtual Agents. (2007). 15-28.

[16] Kleinke, C. 1986. Gaze and Eye Contact: A Research Review. Psychological Bulletin. Vol. 100, No.1. (1986) 78–100.

[17] Lance, B., Marsella, S., Koizumi, D. 2004. Towards Expressive Gaze Manner in Embodied Virtual Agents (2004)

[18] Lance, B., Marsella, S. 2007. Emotionally Expressive Head and Body Movement During Gaze Shifts. Proceedings of Intelligent Virtual Agents. (2007). 72-85.

[19] Lee, S., Badler, J., Badler, N. 2002. Eyes Alive. ACM Transactions on Graphics Vol. 21, No.3. (2002). 637–644.

[20] Leigh, R.J., Zee, D. 2006. The Neurology of Eye Movements, 4th Ed. Oxford University Press, (2006).

[21] Liu, C.K., Hertzmann, A., Popović, Z. 2005. Learning Physics-Based Motion Style with Nonlinear Inverse Optimization. Proceedings of SIGGRAPH, ACM Press, New York (2005)

[22] Mehrabian, A. 1981. Silent Messages: Implicit Communication of Emotions and Attitudes, 2nd edn. Wadsworth Publishing Company (1981).

[23] Mehrabian, A., Ksionzky, S. 1972. Some Determiners of Social Interaction. Sociometry, Vol 35. (1972). 588-609.

[24] de Meijer, M. 1989. The Contribution of General Features of Body Movement to the Attribution of Emotions. Journal of Nonverbal Behavior. Vol 13, No. 4 (1989). 247-268.

[25] Mignault, A., Chaudhuri, A. 2003. The Many Faces of a Neutral Face: Head Tilt and Perception of Dominance and Emotion. Journal of Nonverbal Behavior Vol. 27 No.2. (Summer 2003).

[26] Paterson, H., Pollick, F., Sanford, A. 2001. The Role of Velocity in Affect Discrimination. Proceedings of the 23rd Annual Conference of the Cognitive Science Society (2001).

[27] Pelachaud, C., Bilvi, M. 2003. Modelling Gaze Behavior for Conversational Agents. Proceedings of Intelligent Virtual Agents 2003. LINAI Series, Springer, Heidelberg (2003).

[28] Pelachaud, C., Poggi, I. 2002. Subtleties of facial expressions in embodied agents. The Journal of Visualization and Computer Animation. Vol 13, No. 5. (2002). 301-312.

[29] Rickel, J., Johnson, W.L. 1999. Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control. Applied Artificial Intelligence 13(4-5). (1999). 343–382.

[30] Sobczynski, P. 2004. "Polar Express, The". http://www.efilmcritic.com/review.php?movie=10918&reviewer=389. November 2004.

[31] Stark, L., Zangemeister, W. H., Edwards, J., Grinberg, J., Jones, A., Lehman, S., Lubock, P., Narayan, V., Nystrom, M. 1980. Head Rotation Trajectories Compared with Eye Saccades by Main Sequence Relationships. Investigative Ophthalmology and Visual Science. Vol 19, No. 8. (1980). 986-988.

[32] Thomas, F., Johnston, O. The Illusion of Life: Disney Animation. *Walt Disney Productions.* (1981).

[33] Wallbott, H. G. 1998. Bodily Expression of Emotion. European Journal of Social Psychology. Vol 28 (1998). 879-896.

[34] Witkin, A., Popovic, Z. 1995. Motion Warping. Proceedings of SIGGRAPH, ACM Press, New York (1995).

[35] Zee, D. 1976. Disorders of Eye-Head Coordination. Brooks, B. A., Bajandas, F.J. Eye Movements: ARVO Symposium. Plenium Press. (1976). 9-40.

[36] Zhao, J., Badler, N. 1994. Inverse Kinematics Positioning Using Nonlinear Programming for Highly Articulated Figures. ACM Transactions on Graphics. Vol 13. No. 4. (Oct. 1994). 313-336.

[37] Zhao, L., Badler, N. 2005. Acquiring and Validating Motion Qualities from Live Limb Gestures. Graphical Models. Vol. 67 No. 1. (2005). 1–16.