Final version of paper to appear in JAAMAS 09

Journal of Autonomous Agents and Multiagent Systems manuscript No.
(will be inserted by the editor)

Glances, Glares, and Glowering:

How Should a Virtual Human Express Emotion Through Gaze?

Brent Lance · **Stacy Marsella**

Received: date / Accepted: date

Abstract Gaze is an extremely powerful expressive signal that is used for many purposes, from expressing emotion to regulating human interaction. The use of gaze as a signal has been exploited to strong effect in hand-animated characters, greatly enhancing the believability of the character's simulated life. However, virtual humans animated in real-time have been less successful at using expressive gaze. One reason for this is that a gaze shift towards any specific target can be performed in many different ways, using many different expressive manners of gaze, each of which can potentially imply a different emotional or cognitive internal state. However, there is currently no mapping that describes how a user will attribute these internal states to a virtual character performing a gaze shift in a particular manner. In this paper, we begin to address this by providing the results of an empirical study that explores the mapping between an observer's attribution of emotional state to gaze. The purpose of this mapping is to allow for an interactive virtual human to generate believable gaze shifts that a user will attribute a desired emotional state to. We have generated a set of animations by composing low-level gaze attributes culled from the nonverbal behavior literature. Then, subjects judged the animations displaying these attributes. While the results do not provide a complete mapping between gaze and emotion, they do provide a basis for a generative model of expressive gaze.

Keywords Gaze · Nonverbal Behavior · Emotion · Expression · Procedural · Animation · Posture · Virtual Agent

B. Lance
University of Southern California Institute for Creative Technologies
E-mail: blance@ict.usc.edu

S. Marsella
University of Southern California Information Sciences Institute
E-mail: marsella@isi.edu

1 Introduction

Animated characters in feature films function at a high level of believability, appear to come alive, and successfully engage the film's audience; as do characters in many video games, although arguably to a lesser extent. Unfortunately, the autonomous animated characters developed by the virtual humans community struggle to achieve this goal. Additionally, the animation methods, such as hand-animation and motion capture, used to create the film and video game characters are expensive and time consuming. To use these methods, each physical behavior the virtual human could potentially portray would need to be hand-animated or motion-captured in advance, leading either to long and expensive development cycles, or to limited interaction in dynamic environments. Instead, real-time animation systems that express believable behavior are necessary.

Our specific interest is in gaze behavior, which is expressive not only in terms of where the gaze is directed, but also in the physical manner of a gaze, how it is performed. However, there is currently no mapping that describes how a user will attribute internal states necessary to a lifelike appearance, such as emotion or cognition, to a virtual character performing a gaze shift in a particular manner. In this paper, we begin to address this by providing the results of an empirical study that explores the mapping between an observer's attribution of emotional state to gaze. The purpose of this mapping is to allow for an interactive virtual human to generate believable gaze shifts that a user will attribute a desired emotional state to.

Our previous work on the Expressive Gaze Model (EGM) [14, 15], helps address the difficulty virtual humans have with expressive gaze by providing a means to generate a gaze shift towards an arbitrary target that displays an arbitrary emotion. The EGM is composed of two elements. The first is the Gaze Warping Transformation (GWT) [14], a motion-capture based method for producing emotionally expressive head and torso movement during gaze shifts.

The GWT represents "emotional gaze manner." It is derived from the difference between motion capture data of an emotionally neutral and an emotionally expressive gaze shift, both directed from the same initial posture to the same target. This transformation, when applied to an emotionally neutral gaze shift towards a different, arbitrarily placed target, will modify that neutral gaze shift into one which displays the same expressive behavior as the original emotionally expressive gaze shift used to produce the GWT. This head and torso movement is then integrated with the second component of the EGM, a procedural model of animated eye movement.

There are several benefits to this hybrid approach. By combining an animation method based on motion capture with a procedural animation method, high quality animation that produces gaze shifts to arbitrary targets while displaying desired behavior can be obtained. In addition, the gaze movement is based firmly in human behavior by generating the head and torso movement with motion capture data, and drawing both the model of eye movement - and the integration of eye movement and motion capture - from work in the visual neuroscience literature, such as [17].

However, our goal is to allow virtual humans to produce emotionally expressive gaze shifts with a minimum of motion capture data. The most intuitive way to do this would be to have a Gaze Warping Transformation for each desired emotional state, and then apply that emotional GWT to an emotionally neutral gaze shift directed towards an arbitrary target. While this means that it is no longer necessary to provide the virtual human with a large motion capture library that contains separate gaze shifts for displaying every desired emotion to every possible gaze target, the library would still require a motion capture for every desired emotional state. This may be difficult to obtain, and will still require a potentially large collection of motion captures if we are interested in subtle combinations or varying degrees of emotion.

In this paper, we describe an empirical study that provides an alternative to this intuitive approach by mapping between individual gaze behaviors and emotion. This will allow a further reduction in the size of the motion capture library required by allowing the virtual agent to combine individual gaze behavior components in order to realize the overall expressive quality the agent desires. Thus, the goal of this approach is to obtain a set of low-level gaze behaviors annotated with emotional data that can then be combined according to a model of emotion such that the attribution of emotion to the resulting gaze shift can be predicted. This will allow the generation of a broader array of emotional responses while using a smaller motion library than would a one-to-one method of mapping between physical behavior and emotional attribution.

However, this raises the question, why perform an empirical study at all? Why not just use the mappings between emotion and behavior already reported in the nonverbal behavior literature? Unfortunately, the relationship between behavior and emotion in this literature is still too unclear to simply use it as a mapping. This is due to a number of factors: first, it is not known exactly how behavior expresses emotion. While many behaviors have been shown to be expressive, it is not always clear what is expressed. For example, head angle can express both dominance and pleasure [19].

Second, even if reliable maps existed between emotion and behavior, it may not follow that motion captures of these behaviors would replicate those results. For example, an actor could perform behavior in a subtly idiosyncratic fashion, leading to different interpretations of the behavior. Additionally, 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 exactly the turn should be performed.

For these reasons, our empirical study was performed based on the “reverse engineering” approach of Grammer and Oberzaucher [10]. In this context, “reverse engineering” is used to mean a non-interpretive bottom-up approach where nonverbal behavior expressions are generated through the combination of low-level physical behaviors, and then displayed to subjects who rate the expression on its emotional content. The goal of Grammer and Oberzaucher’s work was to obtain an emotional annotation of Facial Action Coding System (FACS) Action Units (described in [7]) in order to predict the attribution of emotion to a complex facial expression generated from the combination of emotionally annotated Action Units. Specifically, Grammer and Oberzaucher used Poser - a 3d human model animation suite - to generate random facial expressions from the space of all possible combinations of FACS Action Units. Subjects then evaluated the resulting expressions using the circumplex model of emotion, a two-dimensional model of emotion based on the dimensions Valence and Arousal, providing a mapping between the expressive behavior space defined by the FACS Action Units, and the emotional space defined by the circumplex model.

We used a similar process to find a model that describes the mapping between gaze behaviors and the attribution of emotion to gaze shifts displaying those behaviors. For this, we used two different representations of emotion: a set of emotional categories, such as anger or fear; and the Pleasure-Arousal-Dominance dimensional model of emotion [20]. Then we determined a space of possible gazes and the physical manners which they perform. To do this, we have culled a set of low-level, composable gaze behaviors, such as bowing the head during a gaze shift, from the nonverbal behavior literature. We collected motion capture of these behaviors, which we produced GWTs from. We then composed the GWTs of the behaviors and used the resulting movement to drive animated characters. Finally, subjects attributed emotion to the animated characters displaying these behaviors during gaze shifts.

As a result of this reverse engineering study, we were able to demonstrate that composition of these low-level gaze behaviors preserved the PAD dimensional ratings. These results, while promising, do not provide a complete mapping between gaze and emotion. However, they do provide a basis for a generative model of expressive gaze.

While these results have the most application to our GWT-based gaze model, any gaze model with sufficient control over the animation curves used to generate gaze shifts should be able to take advantage of this mapping. In addition, the results also further our basic understanding of how gaze behavior expresses emotion.

2 Related Work

There have been many previous attempts to systematically determine how emotion is attributed to bodily expression in both humans [30], and in virtual characters, such as Coulson's [5] work on the attribution of emotion to rendered images of mannequins in various static postures. More recently, researchers have examined this question with the express intent of applying the answer to animated virtual humans, such as Grammer's [10] work on facial expression, or Shaarani's [29] work on static postures of 3D characters. In this work, we continue to examine this larger question by examining the attribution of emotion to video clips of animated characters performing gaze shifts.

There also have been many implementations of gazing behaviors in real-time applications such as embodied virtual agents. Several of these gaze implementations in virtual characters are based on communicative signals (e.g. [2,24]). Other gaze models have been developed for agents that perform tasks in addition to dialog [9,27]. There are also models of resting gaze, which simulate eye behavior when the eye is not performing any tasks [16,6]. Additionally, there are attention-based models of gaze that perform eye movements based on models of attention and saliency [25,26].

There are several trends which can be seen in these implementations of gaze. First, the models focus on when and where the character looks, not on how the gaze shift occurs. Second, these models, with few exceptions, focus on communicative or task-related gaze behaviors, not on how gaze reveals emotional state.

In addition to the previous research on implementing models of nonverbal gazing behavior, there has been recent work focused on the manipulation of parameters describing the way in which general movement is performed. This concept is referred to as manner or style. This research can provide methods for manipulating the way in which movements are performed, or to obtain the style from one movement and transfer it to another [1,3,11]. This research was inspirational to the development of the Gaze Warping Transformation, but does not deal with the constraints specific to gaze movement, nor does it identify specific styles and their expressive meaning, which is the purpose of this study.

3 Expressive Gaze Model

We used our previous work on gaze to generate the gaze shifts for this study. Our gaze model combines two parts: first, a parameterization called the Gaze Warping Transformation (GWT), that generates emotionally expressive head and torso movement during gaze shifts [14]. The GWT is a set of parameters that transforms an emotionally neutral gaze shift towards a target into an emotionally expressive gaze shift directed at the same target. A small number of GWTs can then produce gazes displaying varying emotional content directed towards arbitrary targets.

The second part [15] is a procedural model of eye movement based on stereotypical eye movements described in the visual neuroscience literature [17]. The procedural eye movement is automatically layered framewise onto the GWT-generated head and torso movement. Emotion is expressed using the GWT, while the procedural eye model ensures realistic motion.

3.1 Gaze Warping Transformation

A Gaze Warping Transformation, or 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 a set of warping 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 [14].

This works by transforming the keyframes of each animation curve. The keyframes of an animation are a subset of that animation's frames, such that the values of the motion curves for intermediate frames are found by interpolating between the keyframes. We select the keyframes for each gaze by aligning it to a "stereotypical" gaze shift with known keyframe locations [15]. The gazes are aligned using the ratio of movement that occurred by each frame to that throughout the entire curve. Key frames may not be evenly spaced in time, but are instead placed where the interpolation will best recreate the values of the intermediate frames.

The result of this is a set of keyframes $x(t)$, defined as a set of value, frame pairs, (x_i, t_i). These keyframes are transformed to those of a new motion $x'(t')$, defined as the set of pairs (x'_i, t'_i) through the use of two functions. The first function, given a frame in the emotional curve t'_i, calculates the location of the corresponding frame t_i in the neutral motion curve. For the GWT, we use the function:

$$t_i = c(t'_i) * (t'_i - t'_{i-1}) \tag{1}$$

where given a frame time in the emotional movement t'_i, the corresponding frame t_i in the neutral movement is determined through a scaling parameter $c(t'_i)$, which scales the time span between two adjacent keyframes. The second function is:

$$x'(t'_i) = x(t_i) + b(t_i) \tag{2}$$

where $b(t_i)$ is a spatial offset parameter that transforms the neutral curve amplitude $x(t_i)$ into the corresponding emotional amplitude $x'(t'_i)$. 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.

Table 1 Classes of Gaze Movement

Gaze Movements
Eye-Only Gaze Shift
Eye-Head Gaze Shift
Eye-Head-Torso Gaze Shift
Head-Only Movement
Head-Torso Movement

As the GWT is based on simple geometric transformations [1,31], the generated animations can move outside the physical limits of a human body. To solve this, we use an inverse kinematics system implemented using nonlinear optimization. This system simulates a rigid skeleton, keeping our animated movement within the limits of the human body [15].

3.2 Procedural Model of Eye Movement

In addition to the GWT, which describes head and torso movement during gaze shifts, we developed an integrated procedural model of eye movement [15]. This model of eye movement is based on the visual neuroscience literature, specifically on research describing the different movements eyes perform during gaze, and the way in which eye movement and head movement are integrated during gaze shifts [17]. It generates several classes of gaze movements (Table 1) using the following building blocks:

Saccades: The saccade is a very rapid, highly-stereotyped eye movement which rotates the eye from its initial position directly to the target. We approximate the main sequence relationship as a linear relation between the saccade amplitude and the number of animation frames the saccade takes to execute. The velocity is implicitly determined by the amplitude and duration.

Vestibulo-Ocular Reflex (VOR): Through the VOR, the eyes rotate within their orbit so that the gaze maintains the same target while the head moves. The VOR produces the Head-Only and Head-Torso movements, and is implemented by counter-rotating the eyes to the head rotation, maintaining the same gaze target.

Combined Eye-Head Movement: This is used to integrate eye movement and head-torso movement, and generates the Eye-Head and Eye-Head-Torso gaze shifts. The position of the eye during this movement is determined by generating a saccade to the target once the head has turned more than 1 away from its starting location. Once the eyes reach the target, the VOR keeps the eyes on target as the head continues to turn.

4 Approach

We performed an empirical study to determine a mapping between a space of gaze behaviors and the emotion that subjects attributed to gaze shifts performing the behaviors. To obtain this mapping, we first selected appropriate emotional models and the space of gaze behaviors to map between. To determine the mapping between a particular gaze and the attribution of emotion to that gaze, we use a “reverse engineering”

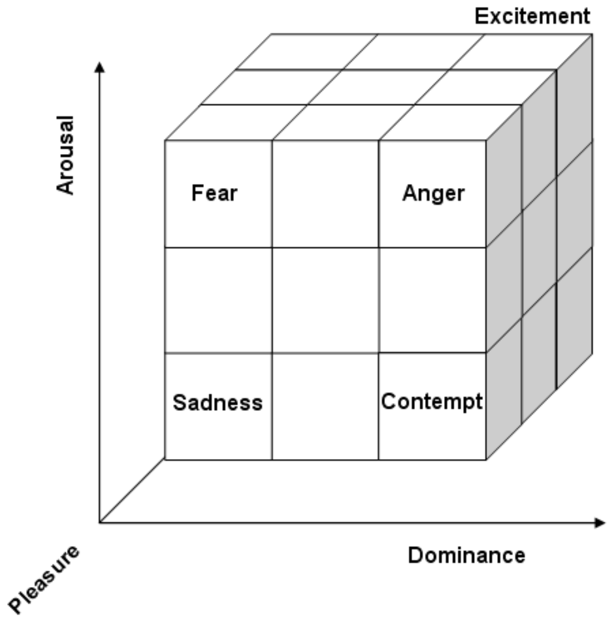

Fig. 1 PAD Dimensional Model of Emotion with Example Emotional Categories

approach [10]. Specifically, we generate all combinations of gaze behavior allowed by our space of gaze behaviors, and collect data of subjects attributing emotion to the resulting gaze shifts.

4.1 Structure of Model

Selected Emotion Model. There are many potential models of emotion we could have mapped to the gaze behaviors. For this study, we have selected two: the first is the Pleasure-Arousal-Dominance or PAD model [20]; a model of emotion that views emotion as a space described with a three dimensions: pleasure / displeasure, arousal / non-arousal, and dominance / submissiveness. While there are many possible alternative emotional models, the PAD dimensional model is composed of a manageable number of dimensions, each of which have a background of research describing nonverbal behaviors associated with them (for example, see [23, 19]).

We are also mapping gaze behaviors to a set of intuitive emotional categories. These categories of emotion, such as anger or sadness, can be represented in the PAD model by subregions in the space defined by the emotional dimensions. For example, anger can be defined as negative valence, high arousal, and high dominance, while fear can be defined as negative valence, high arousal, and low dominance, and sadness can occupy a region where valence is negative, and arousal and dominance are both low. A visual depiction of the emotional dimensions and one possible division into categories is shown in Figure 1. Rather than using an existing categorical model, this categorization was derived from observer responses to the animations.

Table 2 Gaze Behaviors

Emotional Gaze Behaviors
Head Raised
Head Bowed
Faster Velocity
Slower Velocity
Torso Raised
Torso Bowed

Selected Gaze Behavior. In order to use this reverse engineering approach, a set of gaze behaviors to reproduce in animations is required. Therefore, we also had to determine a space of gaze behaviors, due to the lack of a descriptive set of gaze behaviors analogous to the FACS system used by Grammer and Oberzaucher [10]. We identified a set of emotional behavior attributes from the psychology and arts literature that are likely to reveal emotional state. This set of behavior attributes can be seen in Table 2.

Because the emphasis of this work is on generating emotional behaviors that can be correctly recognized by observers and associating these behaviors with emotions using a bottom-up behavior-based method, the focus for selecting behaviors was on the identification of a space of behaviors that can be used to express emotion in gaze, not on what emotional states these behaviors actually display.

As a result, these guidelines are simplifications of the actual literature [13]. The literature indicates that vertical head orientation will affect the display of dominance [19], that the perception of arousal is strongly related to velocity [14], and that vertical posture of the body will display emotional pleasure [28]. While there are many alternative gaze behaviors that could also be modeled using the GWT, such as subtle variations in dynamics, or wider variations on posture, this limited set provides a starting point for this research.

4.2 Motion Capture Collection

For the head and torso behaviors, we asked the actor to perform “raised,” “neutral,” and “bowed” versions of the behavior, and collected data from the resulting movement. We also collected “fast,” “neutral,” and “slow” velocity movements. However, the “raised” torso posture was indistinguishable from the “neutral” torso posture, due to the limitations of the motion tracking system we used, resulting in the set of physical behaviors shown in Table 3. All captured gaze shifts consisted of the desired behavior being displayed in a gaze aversion that started gazing straight ahead in a neutral position and posture, and ended gazing 30 degrees to the right displaying the intended gaze behavior. From this motion data, we produced eight behavior GWTs, one for each behavior listed in Table 3.

Table 3 Discretization of Gaze Behaviors

Behavior Dimension	Possible Values
Head Posture	Raised, Neutral, Bowed
Torso Posture	Neutral, Bowed
Movement Velocity	Fast, Neutral, Slow

Table 4 Combinations of Gaze Behaviors

Head Posture	Torso Posture	Movement Velocity	Label
Raised	Neutral	Fast	RNF
Raised	Neutral	Neutral	RNN
Raised	Neutral	Slow	RNS
Neutral	Neutral	Fast	NNF
Neutral	Neutral	Neutral	NNN
Neutral	Neutral	Slow	NNS
Neutral	Bowed	Fast	NBF
Neutral	Bowed	Neutral	NBN
Neutral	Bowed	Slow	NBS
Bowed	Neutral	Fast	BNF
Bowed	Neutral	Neutral	BNN
Bowed	Neutral	Slow	BNS
Bowed	Bowed	Fast	BBF
Bowed	Bowed	Neutral	BBN
Bowed	Bowed	Slow	BBS

We also collected motion capture and produced GWTs of the different classes of gaze movement (Table 1). The classes of gaze movement were captured both as aversions that began gazing straight ahead and ended gazing 30 to the right, and as attractions that began 30 degrees to the right and ended gazing straight ahead. This resulted in 10 GWTs - one aversive and one attractive gaze movement for each of the different classes of gaze movement in Table 1.

4.3 Animation Generation

From these 8 GWTs representing the discretized physical behaviors (Table 3) and 10 GWTs representing the various classes of gaze movement (Table 1), we generated 150 animations for use in our empirical bottom-up study. To do this, we first composed the gaze behaviors in all possible ways, leaving out combinations of a raised head with bowed torso due to the physical implausibility of the behavior, resulting in 15 total behavior combinations (shown in Table 4). These behavior combinations provide the emotionally expressive content of the generated gaze shifts. Then, these combined gaze behaviors were composed with the 10 gaze movement GWTs from Table 1, that describe how the eye movement is related to the head movement, resulting in 150 GWTs that demonstrated both a wide range of expressive behavior, and a wide range of eye-head movement relationships. Finally, to generate the animations, we applied these 150 GWTs to neutral gaze shifts, with the resulting output rendered using Maya. These animations can be seen at:

`http://people.ict.usc.edu/~blance/AnimationGallery/AnimationGallery.html`

4.4 Category Formation

In order to determine the categories for our primary experiment and obtain a picture of how well the animated gaze behaviors covered the emotional space defined by the emotion models, we performed a preliminary category formation study.

Table 5 Emotional Categories

Emotional Categories	
Anger	Contempt
Disbelief	Excitement
Fear	Flirtatious
Guilt	Sadness
Secretive	Surprise

Approach. 31 subjects each viewed 20 animations randomly selected from the set of 150 animations with no duplicates, giving us 620 views, or approximately 4 per animation, and provided an open-ended written response to the question “What emotional state is the character displaying?” We then categorized the affective responses based on the hierarchical model of emotion described in [21].

Results. We used the hierarchical model as a sorting guideline, to divide the individual responses into ten categories (Table 5); for example categorizing “expression of contempt” as Contempt, or “terrified” as Fear. However, we also utilized additional categories that were common in the subject’s open ended responses but not described by the hierarchical model. After categorizing the responses, we then selected categories where at least one video had 50% of the subjects rate it with that category.

We then discarded those categories that were related to attention, discarding responses such as “change in attention,” “displaying strong interest,” and “distracted.” Since every gaze inherently indicates interest, attention, or distraction, we were concerned that these types of categories would become “catch-alls,” and draw a disproportionate number of responses during the forced-choice selection, while providing no relevant information. Finally, we discarded the responses indicating “uncertainty,” as we were concerned that it would be applied when the subject was uncertain of the character’s state, not when the character was displaying uncertainty.

4.5 Emotional Attribution Experiment

After selecting the low-level behaviors, generating the animations, and setting the emotional categories, we performed the empirical study. The animations were placed online, and subjects rated the animation in two ways: first by selecting the emotional category (Table 5) that most closely approximated the emotion that they perceived in the animation, and second by locating the animation’s perceived emotion along the emotional dimensions of the PAD model.

Table 6 Emotional Dimension Rating Scales

Emotional Dimension	Rating Statement
High Dominance	The character is dominant.
Low Dominance	The character is submissive.
High Arousal	The character is agitated.
Low Arousal	The character is relaxed.
High Valence	The character is pleased.
Low Valence	The character is displeased.

Subjects rated the animation's location within the PAD model by using five-point scales to indicate their agreement with two statements representing each dimension, seen in Table 6. The scale items were 1 = Strongly Disagree, 2 = Disagree, 3 = N/A, 4 = Agree, 5 = Strongly Agree. Both emotional categories and rating statements were displayed in random order for each animation. One hundred people each rated fifteen unique, randomly selected animations, resulting in ten ratings for each of the 150 animations.

5 Results

We uncovered the mapping between emotion models and physical behaviors in order to answer the following questions:

1. How did the PAD ratings relate to the low-level gaze behaviors in Table 3?
2. Can these low-level gaze behaviors be composed within the PAD dimensions?
3. Can low-level gaze behaviors be combined across PAD dimensions into emotional categories?

We had originally intended to find how the 150 individual animations varied across emotional state, but ten ratings per animation was too few to perform a reliable statistical analysis. Instead, we combined gazes across the classes of gaze movement (Table 1), giving us 50 ratings for each of the 15 combinations of gaze behaviors (Table 4).

5.1 Dimensional Results

How reliable were the dimensional ratings scales? Before exploring the dimensional results, we tested how well our dimensional rating scales measured the emotional dimensions they were intended to by calculating the correlation and Cronbach's Alpha between each pair of rating scales from Table 6.

The Pleased and inverted Displeased scales performed well. The correlation between them was 0.615, and the standardized Alpha score indicating scale reliability was high, with $\alpha = 0.7610$, ($\alpha > 0.7$ is considered a reliable scale). Dominant and inverted Submission also did well, with a correlation of 0.6649, and a high Alpha ($\alpha = 0.7987$). Therefore, we averaged Pleased and inverted Displeased into one Pleasure scale, and combined Dominant and inverted Submission into one Dominance scale. Correlations between the Dominance and Pleasure scales were low, (0.1569), indicating little overlap.

However using the ratings of Relaxed and Agitated as a scale for Arousal was less reliable, as both correlation (0.3745) and Alpha ($\alpha = 0.5449$) were low. In addition, correlations between Relaxed and Pleased (0.5353) and between Agitated and Displeased (0.4889) were higher than between Relaxed and Agitated. For the remainder of this paper, we will be using the two scales separately as Relaxed and Agitated. As we used 5-point scales, but only animated 3-point scales of physical behavior, we condensed the collected data into 3-point scales by combining "Strongly Disagree" and "Disagree", as well as "Strongly Agree" and "Agree", leaving Neutral ratings unchanged.

How did the PAD dimension ratings relate to the low-level gaze behaviors in Table 3? A series of MANOVAs (multivariate analysis of variance) and post-hoc tests determined whether or not the mean emotion dimensions ratings differed across the low

Table 7 Significant Relationships between PAD Dimension and Individual Gaze Behaviors

Emotional Dimension	Head	Torso	Velocity
High Dominance	Raised	Bowed	Fast
Low Dominance	Bowed	Neutral	Non-Fast
Relaxed		Bowed	
Agitated	Non-Bowed		Fast
High Pleasure	Neutral	Bowed	
Low Pleasure	Non-Neutral	Neutral	

level behaviors found in Table 3. Four MANOVAs were performed, each with one dimension (Pleasure, Agitation, Relaxation, or Dominance) as the dependent variable, and Head Orientation, Torso Orientation, Velocity, and Subject as the independent variables, while testing for second degree factorial interaction between the independent variables. Table 7 provides the results of this analysis, and each row in the table provides alternative methods for signaling that emotional dimension.

As shown in Rows 1 and 2 of Table 7, the ratings for Dominance attributed to gaze shifts significantly differed across differing Head Orientation, Torso Orientation, and Velocities. The MANOVA results for Dominance showed significant effects ($N = 1500, DF = 18, F = 14.5110, p < .001$) for head orientation ($F = 24.0776, p < .001$), torso orientation ($F = 82.5508, p < .001$), and velocity ($F = 7.3838, p < .001$), with a significant interaction between head and torso orientation ($F = 6.4689, p < .05$).

In addition, Rows 1 and 2 show the results of the post-hoc tests. These tests showed clear differences between group means, with higher Dominance ratings corresponding to raised head gaze shifts, and lower Dominance ratings corresponding to bowed head shifts. In addition, the post-hoc tests revealed that a bowed posture was rated as higher Dominance, and a neutral posture was rated for lower Dominance. Finally, the Dominance rating for fast movements was higher than that for slow or for neutral movements (all results significant to $p < .01$).

In contrast, Row 3 displays the post-hoc tests results showing that the Relaxed dimension only significantly differs across torso orientation, with gaze shifts displaying a bowed torso drawing significantly higher Relaxed ratings from subjects ($p < .01$). However, MANOVA results showed significant differences ($N = 1500, DF = 18, F = 1.8892, p < .05$) across both the torso orientation ($F = 11.4132, p < .001$) and the velocity ($F = 3.7849, p < .05$), with a significant interaction effect between torso and velocity ($F = 3.6755, p < .05$). The post-hoc tests did not reveal useful information about the velocity, indicating that the significant difference found by the MANOVA was likely related to the interaction between torso and velocity. Row 4 shows the post-hoc test results that raised and neutral head orientations were rated as significantly more Agitated than bowed head orientation, and that the Agitated rating for high velocity was higher than that for slow or neutral ($p < .05$). The MANOVA for Agitation found significant differences ($N = 1500, DF = 18, F = 4.5978, p < .001$) across the head orientation ($F = 19.6129, p < .001$), the velocity ($F = 6.0387, p < .01$), and the subject ($F = 17.1201, p < .001$), and a significant interaction effect between the head and the velocity ($F = 7.1696, p < .05$).

Finally, post-hoc tests ($p < .01$) revealed that the Pleasure rating for a neutral head orientation was significantly higher than those for bowed and raised head orientations, and that a bowed posture received significantly higher Pleasure ratings than a neutral posture, as shown in Rows 5 and 6. The MANOVA showed that Pleasure significantly

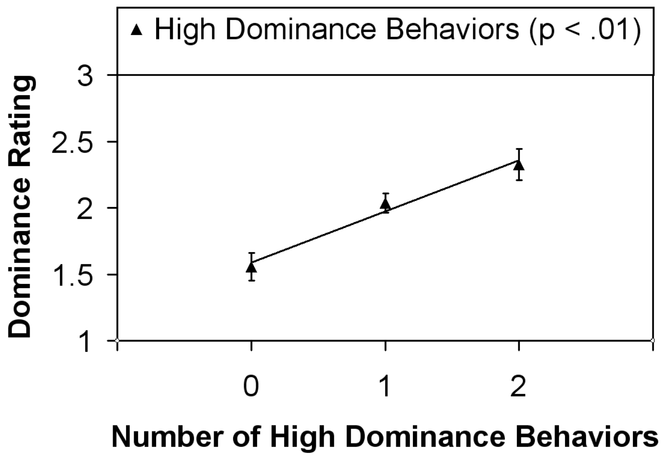

Fig. 2 Mean Dominance Rating vs. Number of High Dominance Behaviors

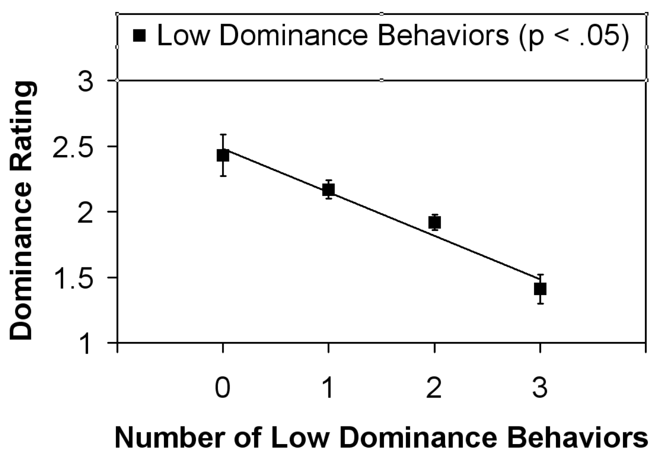

Fig. 3 Mean Dominance Rating vs. Number of Low Dominance Behaviors

differed ($N = 1500, DF = 18, F = 5.9261, p < .001$) across both the vertical orientation of the head ($F = 6.5836, p < .05$) and the torso ($F = 77.5703, p < .001$), with no significant interaction effects. The ratings for Pleasure also differed significantly ($F = 4.0601, p < .05$) across subject.

Can these low-level gaze behaviors be composed within the PAD dimensions? In order to determine whether or not the low-level behaviors can be composed within individual PAD dimensions, a second analysis tested whether or not gaze shifts displaying different numbers of behaviors significantly related to a specific emotional dimension would have different ratings for that dimension attributed to them. For example, does the mean Dominance rating of a gaze shift that displays one Low Dominance behavior significantly differ from the mean Dominance rating of a gaze shift that displays two Low

Dominance behaviors? In turn, do these ratings significantly differ from a gaze shift that displays 3 Low Dominance behaviors? For this analysis, four MANOVAs were

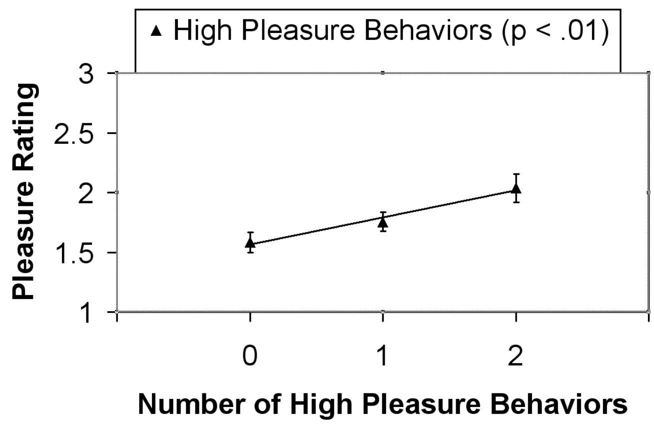

Fig. 4 Mean Dominance Rating vs. Number of High Pleasure Behaviors

Fig. 5 Mean Dominance Rating vs. Number of Low Pleasure Behaviors

performed. Each MANOVA used an emotional dimension (Dominance, Relaxed, Agitated, and Pleasure) as the dependent variable, while the independent variables were the number of behaviors associated with that emotional dimension, and the subject. Thus, one MANOVA had Dominance as a dependent variable, while the independent variables were the number of low dominance behaviors, the number of high dominance behaviors, and the subject. The results of this analysis clearly showed that mean attributed ratings for an emotional dimension increased as the number of gaze behaviors associated with that emotional dimension increased, as seen in Figures 2 through 6. This indicates that physical gaze behaviors, when composed within PAD dimensions will be rated as predicted by the composed behaviors.

The specific results for Dominance show that there were significant differences ($N = 1500, DF = 6, F = 32.2426, p < .01$) across the number of both Low ($F = 14.1668, p < .001$) and High Dominance ($F = 26.3914, p < .001$) behaviors displayed in a gaze shift, and a significant ($F = 6.9287, p < .01$) interaction effect between Low and High Dominance. Post-hoc tests showed that as the number of High Dominance gaze behaviors displayed in a gaze shift increased, the mean rating of Dominance for that gaze shift significantly increased ($p < .01$) as well (Figure 2). In contrast, as the number of Low Dominance behaviors increased, the mean rating of Dominance for that gaze shift significantly decreased (Figure 3). Note that at most two High Dominance behaviors could be composed, due to the incompatibility between a raised head and a bowed torso.

The mean Pleasure rating showed significant differences across the number of both Low and High Pleasure behaviors displayed in a gaze shift ($N = 1500, DF = 3, F = 22.9619, p < .001$), although there were also significant differences across viewers ($F = 4.8669, p < .05$), and no interaction effects. Subsequent post-hoc tests showed that mean ratings of Pleasure significantly increased ($p < .01$) as the number of High Pleasure behaviors displayed in a gaze shift increased (Figure 4); and that mean ratings of Pleasure significantly decreased ($p < .01$) as the number of Low Pleasure behaviors increased (Figure 5).

The MANOVA for Agitated revealed that there were again significant differences across the number of behaviors ($N = 1500, DF = 3, F = 18.3058, p < .001$) displayed in a shift, but also showed significant differences across viewers ($F = 20.5002, p < .001$), although there were no interaction effects. The post-hoc tests demonstrated that gaze shifts with zero or one Agitated behaviors were rated as significantly less Agitated than those shifts with two Agitated behaviors ($p < .01$), although the difference between zero and one behaviors was not significant (see Figure 6).
As the relaxed dimension only had one behavior associated with it, no further testing was performed.

5.2 Categorical Results

Can low-level gaze behaviors be combined across PAD dimensions into emotional categories? To answer this question, we generated a cross tabulation of the 15 combinations of gaze behaviors (Table 4) against the emotional categories (Table 5), and used Pearson's chi squared (χ^2) test to examine relationships in the data. We then performed further tests on the residuals to determine which had significant differences.

Results of this analysis can be seen in Table 8. The χ^2 test showed that gaze combinations and emotional categories were not randomly related ($N = 1500, DF = 126, \chi^2 = 775.817, p < .01$). The table rows show behavior combinations with a significant number ($p < .05$) of ratings for that emotional category. For example, Contempt was significantly associated with the gaze behavior combination of Raised Head, Neutral Torso, and Neutral Velocity, while Excitement was significantly associated with the combination of a Neutral Head, Bowed Torso, and Fast Velocity.

While only 5 of the 15 gaze behavior combinations from Table 4 had significant associations to emotional categories, it was clear through examination of the residuals

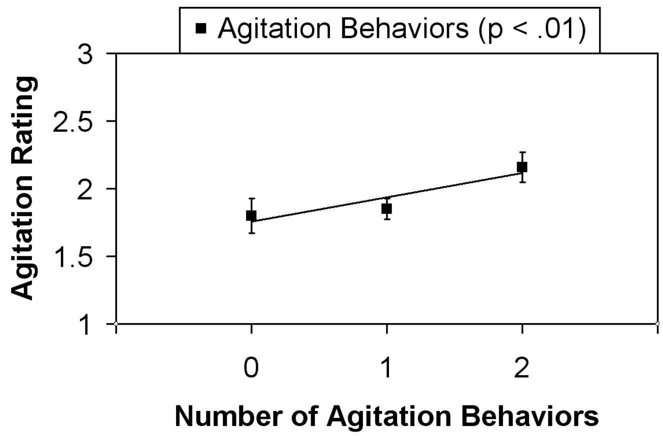

Fig. 6 Mean Dominance Rating vs. Number of Agitation Behaviors

that further study of the relationship between the emotional categories and the low-level behaviors from Table 3 could be useful. For example, while no individual gaze behavior combination was rated significantly high for Flirtatious, all gaze shifts with the bowed head behavior had more Flirtatious ratings than did the gaze shifts without bowed head. To examine this, we generated crosstabs and performed additional χ^2 tests to determine which individual gaze behaviors were significantly correlated against the emotional categories (Table 9).

We found significant interactions between head vertical orientation and emotional categories, $(N = 1500, DF = 18, \chi^2 = 329.47, p < .001)$. Testing the residuals showed that the Contempt category was more likely $(\chi^2 = 70.35, p < .05)$ to be attributed to a gaze shift with the head raised behavior, while Flirtatious $(\chi^2 = 73.41, p < .01)$, Guilt $(\chi^2 = 81.33, p < .01)$, and Sadness $(\chi^2 = 42.51, p < .01)$ were all more likely to be attributed to bowed head gaze shifts. Finally, Surprise was significantly less likely $(\chi^2 = 55.30, p < .01)$ to be attributed to bowed head gazes. Anger, Disbelief, Excitement, Fear, and Secretive did not relate to head vertical orientation significantly.

Torso posture was not randomly related to emotional category $(N = 1500, DF = 9, \chi^2 = 187.49, p < .001)$. Excitement was more likely to have a bowed torso $(\chi^2 = 62.94, p < .01)$, while Contempt $(\chi^2 = 24.24, p < .05)$, Fear $(\chi^2 = 29.19, p < .01)$, and Guilt $(\chi^2 = 19.88, p < .01)$ were attributed more often to neutral torso animations.

We also found no strong relationships between the emotional categories and the velocity of the gaze using a crosstab of emotions by velocity. While, the chi squared test showed that emotional category and velocity are not randomly related $(N = 1500, DF = 18, \chi^2 = 42.36, p < .001)$, upon examination of the residuals, no emotional categories significantly differed across velocities.

6 Discussion

The goal of this reverse engineering study was to attribute emotional state to low-level gaze behaviors, and then test whether or not gaze shifts generated by composing these behaviors would have similar emotions attributed to them. The results are highly

Table 8 Emotional Categories and Significantly Related Behavior Combinations

Emotional Categories	Significantly Related Behavior Combinations
Contempt	Head Raised, Torso Neutral, Velocity Neutral
Excitement	Head Neutral, Torso Bowed, Velocity Fast
Fear	Head Neutral, Torso Neutral, Velocity Neutral
	Head Neutral, Torso Neutral, Velocity Slow
Guilt	Head Bowed, Torso Neutral, Velocity Neutral
	Head Bowed, Torso Neutral, Velocity Slow
Sadness	Head Bowed, Torso Neutral, Velocity Fast
	Head Bowed, Torso Neutral, Velocity Neutral

Table 9 Significant Relationships between Emotional Categories and Individual Behaviors

Emotional Category	Head	Torso
Contempt	Raised	Neutral
Excitement		Bowed
Fear		Neutral
Flirtatious	Bowed	
Guilt	Bowed	Neutral
Sadness	Bowed	
Surprise	Neutral	

promising. The foremost result is that individual gaze behaviors can be associated with individual dimensions in the Pleasure-Arousal-Dominance space (Table 7).

Once this is done, the rating along a single PAD dimension that a subject will attribute to gaze shifts generated through the composition of these individual gaze behaviors can be predicted. For example, a gaze shift containing more behaviors associated with a higher mean Dominance rating will be much more likely to be viewed as a High Dominance gaze shift (see Figures 2 through 6).

In addition, this study demonstrates that limited composition across PAD dimensions is possible, as shown in Table 10. This table shows that the emotional categories of Contempt, Excitement, Guilt, and Sadness are significantly related to the same behavior combinations that are also significantly related to the location of the categorical emotion when mapped into PAD space (mapping based on [12] and [8]).

In Row 1 of Table 10, we provide the result that Contempt is significantly related with the behavior combination of a raised head, a neutral torso, and a neutral velocity (Table 8). The raised head is consistent with a High Dominance (+D), the neutral torso is consistent with Low Pleasure (-P), and the non-fast velocity is consistent with a Low Agitation (-A) (Table 7). Jiang [12] maps Disdainful, an emotional label very similar to Contempt into the -P -A +D quadrant of the PAD model. Thus, by combining the High Dominance (+D), Low Pleasure (-P) and Low Arousal (-A) behaviors into a gaze shift, subjects will attribute the emotional state of Contempt to the resulting gaze shift. Similarly, in Row 2, Excitement is significantly related to the behavior combination of neutral head, bowed torso, and fast velocity (Table 8), which are associated with High Dominance, High Pleasure and High Arousal, respectively (Table 7). Excitement can be mapped into the PAD quadrant of Exuberant (+P +A +D) [20,8].

Row 3 shows Guilt, which is very similar to the emotional label Remorse that is mapped to (-P -A -D) by Jiang [12] and to (-P +A -D) by Gebhard [8]. Guilt was significantly associated with two behavior combinations: bowed head, neutral torso,

Table 10 Composition of Behaviors Across PAD Dimensions into Emotional Categories

Emotional Categories	PAD Rating of Categories	Significantly Related Behavior Combinations and PAD Rating of Individual Behaviors		
Contempt	-P-A+D	Head Raised	Torso Neutral	Velocity Neutral
		+D	-P	-A
Excitement	+P+A+D	Head Neutral	Torso Bowed	Velocity Fast
		+D	+P	+A
Guilt	-P+A-D/	Head Bowed	Torso Neutral	Velocity Neutral
	-P-A-D	Head Bowed	Torso Neutral	Velocity Slow
		-D	-P	-A
Sadness	-P-A-D	Head Bowed	Torso Neutral	Velocity Fast
		-D	-P	+A
		Head Bowed	Torso Neutral	Velocity Neutral
		-D	-P	-A

neutral velocity, and bowed head, neutral torso, and slow velocity (Table 8). These are in turn significantly associated with Low Dominance, Low Pleasure, and Low Arousal (Table 7). Finally, Row 4 shows that Sadness can be mapped into the PAD space as negative Pleasure (-P), negative Arousal (-A), and negative Dominance (-D) [12], and is significantly more likely to be attributed to gaze shifts with a bowed head, a neutral torso, and a neutral velocity (Table 8). In Table 7, it is shown that bowed head and a neutral torso are significantly associated with negative Dominance and negative Pleasure, that neutral velocity are not associated with Agitation.

7 Conclusion

In this paper, we have provided the results of a reverse engineering study resulting in a preliminary mapping between gaze behaviors and emotional states that could be used with a variety of gaze or emotion models. In addition, we have shown that combining low-level behaviors associated with emotional dimensions in accordance with those dimensions generates a gaze shift that subjects attribute the combined emotional state to. These results, while promising, do not provide a complete mapping between gaze and emotion. However, this study demonstrates the utility of the GWT as a nonverbal behavior research tool, and points towards several directions for future research.

Many of the relationships between behavior and emotion found in this study are consistent with previous research. For example, Wallbott [30] found that shame was associated with a bowed head, similar to these findings for guilt; and Coulson [5] showed that upwards tilted head with a neutral torso was associated with disgust, similar to these contempt findings.

However, several of these findings conflict with earlier research as well. Coulson's [5] subjects attributed anger and fear to postures with the torso bowed and the head, while this study found that excitement was attributed to a bowed torso and upright head. In addition, de Meijer [18] found that trunk bowing was strongly correlated with negative affect, while an upright trunk was associated with positive affect, which is the reverse of these findings. Finally, this work does not replicate our previous results showing velocity being highly associated with arousal [14].

There are, however, several possible explanations for these differences. One explanation for this is that the "bowed" torso movement used in this work is not viewed as a bowed movement per se by the subjects. Instead, it is perhaps viewed as an "approach" behavior, where the character is moving towards an object or another person. For example, Carney [4] demonstrated that "leans forward towards other" was perceived as a dominant nonverbal behavior. Another possibility is that the dynamics of these movements changed the perception of the movement when compared to Coulson's [5] static postures, or de Meijer's [18] movements.

The lack of agreement with our own prior results [14] is somewhat more disconcerting. The problem here is twofold. First, the rating scales used for the Arousal dimension did not perform very well. Second, the differences between the low, medium, and high velocities in the videos used for this experiment were much lower than those we used previously [14]. These factors likely account for the discrepancy.

It should also be pointed out that several strong assumptions were made during the course of this study, and that improved mappings between emotion and behavior may be obtained by relaxing these assumptions. The first assumption is that the mapping between categorical emotions and PAD space is correct, or even that a consistent mapping exists. As such, improved mappings between emotion and behavior could be obtained by utilizing an alternate mapping between categorical emotions and dimensional emotions, or by forgoing a categorical model of emotion entirely.

Another assumption made is that gaze behaviors affect the attribution of emotion independently from each other. In other words, it is assumed that adding any specific behavior to a gaze shift will affect the attribution of emotion to that gaze shift in the same way regardless of any other behaviors the gaze shift portrays. Improved emotion/behavior mappings may be obtained by relaxing this assumption and examining how each gaze behavior affects the attribution of emotion to the other gaze behaviors displayed in a gaze shift, as well as for the gaze shift as a whole.

Finally, a rigorously defined set of low-level behaviors used for expressive gaze would be very valuable to this type of research. While a space of gaze behaviors was determined for this study, there are other possible ways to structure this space that may provide better results. As Pasch and Poppe [22] note, there is no agreed-upon physical parameterization for static posture studies, much less the more complex problem of movement during gaze.

Acknowledgements Dr. Bosco Tjan, Dr. Skip Rizzo, Dr. Sinhwa Kang, Mei Si, and Tiffany Cole provided invaluable 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.

References

1. Amaya, K., A. Bruderlin, and T. Calvert. 1996. Emotion from motion. In *Proceedings of the conference on Graphics interface '96* 222-229. Canadian Information Processing Society Toronto, Ont., Canada, Canada.
2. Bickmore, T., and J. Cassell. 2004. Social dialogue with embodied conversational agents. *Natural, Intelligent and Effective Interaction with Multimodal Dialogue Systems*. New York: Kluwer Academic: 2005-05.
3. Brand, M., and A. Hertzmann. 2000. Style machines. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, 183-192.
4. Carney, D. R., J. A. Hall, and L. S. LeBeau. 2005. Beliefs about the nonverbal expression of social power. *Journal of Nonverbal Behavior* 29, no. 2: 105-123.

5. Coulson, M. 2004. Attributing Emotion to Static Body Postures: Recognition Accuracy, Confusions, and Viewpoint Dependence. *Journal of Nonverbal Behavior* 28, no. 2: 117-139.
6. Deng, Z., J. P. Lewis, and U. Neumann. 2005. Automated Eye Motion Using Texture Synthesis. *IEEE Computer Graphics and Applications*: 24-30.
7. Ekman, P., and E. L. Rosenberg. 2005. *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. Oxford University Press, USA.
8. Gebhard, P. 2005. ALMA: A Layered Model of Affect. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, 29-36. ACM New York, NY, USA.
9. Gillies, M. F. P., and N. A. Dodgson. 2002. Eye Movements and Attention for Behavioural Animation. *The Journal of Visualization and Computer Animation* 13, no. 5: 287-300.
10. Grammer, K., and E. Oberzaucher. The Reconstruction of Facial Expressions in Embodied Systems. Mitteilungen, *ZiF*.
11. Hsu, E., K. Pulli, and J. Popovic. 2005. Style Translation for Human Motion. In *International Conference on Computer Graphics and Interactive Techniques*, 1082-1089. ACM Press New York, NY, USA.
12. Jiang, H. 2007. *From Rational to Emotional Agents*. University of South Carolina.
13. Kleinke, C. 1986. Gaze and Eye Contact: A Research Review. *Psychological Bulletin* 100, no. 1: 78-100.
14. Lance, B., and S. C. Marsella. 2007. Emotionally Expressive Head and Body Movement During Gaze Shifts. *Lecture Notes in Computer Science 4722*: 72.
15. Lance, B. J., and S. C. Marsella. 2008. A Model of Gaze for the Purpose of Emotional Expression in Virtual Embodied Agents. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1*, 199-206. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC.
16. Lee, S. P., J. B. Badler, and N. I. Badler. 2002. Eyes Alive. *ACM Transactions on Graphics (TOG)* 21, no. 3: 637-644.
17. Leigh, R. J., and D. S. Zee. 2006. *The Neurology of Eye Movements*. New York: Oxford University Press.
18. de Meijer, M. 1989. The Contribution of General Features of Body Movement to the Attribution of Emotions. *Journal of Nonverbal Behavior* 13, no. 4: 247-268.
19. Mignault, A., and A. Chaudhuri. 2003. The Many Faces of a Neutral Face: Head Tilt and Perception of Dominance and Emotion. *Journal of Nonverbal Behavior* 27, no. 2: 111-132.
20. Mehrabian, A. 1981. *Silent Messages: Implicit Communication of Emotions and Attitudes*. Wadsworth Pub Co.
21. Metts, S., and J. W. Bowers. 1994. Emotion in Interpersonal Communication. In *Handbook of Interpersonal Communication*, 508-541. 2nd ed. Sage Publications.
22. Pasch, M., and R. Poppe. 2007. Person or Puppet? The Role of Stimulus Realism in Attributing Emotion to Static Body Postures. *Lecture Notes in Computer Science* 4738: 83.
23. Paterson, H. M., F. E. Pollick, and A. J. Sanford. 2001. The Role of Velocity in Affect Discrimination. In *Twenty-Third Annual Conference of the Cognitive Science Society*, Edinburgh.
24. Pelachaud, C., and M. Bilvi. 2003. Modelling Gaze Behavior for Conversational Agents. In *Intelligent Virtual Agents: 4th International Workshop, IVA 2003*. Springer Verlag.
25. Peters, C., C. Pelachaud, E. Bevacqua, M. Mancini, and I. Poggi. 2005. A Model of Attention and Interest Using Gaze Behavior. *Lecture Notes in Computer Science* 3661: 229.
26. Picot, A., G. Bailly, F. Elisei, and S. Raidt. 2007. Scrutinizing Natural Scenes: Controlling the Gaze of an Embodied Conversational Agent. *Lecture Notes in Computer Science* 4722: 272.
27. Rickel, J., and W. L. Johnson. 1999. Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control. *Applied Artificial Intelligence* 13, no. 4-5: 343-382.
28. Schouwstra, SJ, and J. Hoogstraten. 1995. Head Position and Spinal Position as Determinants of Perceived Emotional State. *Perceptual and Motor Skills* 81, no. 2: 673-674.
29. Shaarani, A. S., and D. M. Romano. 2008. The Intensity of Perceived Emotions in 3D Virtual Humans. *In Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 3*, 1261-1264. International Foundation for Autonomous Agents and Multiagent Systems Richland, SC.
30. Wallbott, HG. 1998. Bodily Expression of Emotion. *European Journal of Social Psychology* 28: 879-896.

31. Witkin, A., and Z. Popovic. 1995. Motion Warping. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, 105-108. ACM New York, NY, USA.