Chapter 4
Eye Movements, Saccades, and Multiparty Conversations

Erdan Gu, Sooha Park Lee, Jeremy B. Badler, and Norman I. Badler

4.1 Introduction

In describing for artists the role of eyes, Faigin [20] illustrates that downcast eyes, upraised eyes, eyes looking sideways, and even out-of-focus eyes are all suggestive of states of mind. Given that eyes are a window into the mind, we propose a new approach for synthesizing the kinematic characteristics of the eye: the spatiotemporal trajectories of saccadic eye movement. "*Saccadic eye movements take their name from the French 'saccade', meaning 'jerk', and connoting a discontinuous, stepwise manner of movement as opposed to a fluent, continuous one. The name very appropriately describes the phenomenological aspect of eye movement*" [4].

We present a statistical eye movement model based on both empirical studies of saccades and acquired eye movement data. There are three strong motivations for our work. First, for animations containing close-up views of the face, natural-looking eye movements are desirable. Second, traditionally it is hard for an animator to obtain accurate human eye movement data. Third, the animation community appears to have had no models for saccadic eye movement models that are easily adopted for speaking or listening faces. We apply the eye model to conversational agents in which gaze direction and role are modeled on saccades during talking, listening, and "thinking" as well as on the social aspects of interaction behaviors such as turn-taking and feedback signals. A preliminary eye saccade model is the basis for the present work [28].

As computer animation techniques mature, there has been considerable interest in the construction and animation of human facial models. Applications include such diverse areas as advertising, film production, game design, teleconferencing, social agents and avatars, and virtual reality. To build a realistic face model, many factors including modeling of face geometry, simulation of facial muscle behavior, lip synchronization, and texture synthesis have been considered. Several early researchers [25, 32, 37, 43] were among those who proposed various methods to simulate facial shape and muscle behavior. A number of investigators have recently emphasized building more realistic face models [8,21,30,36]. Others have suggested automatic methods of building varied geometric models of human faces [7, 16, 29]. Motion capture methods can be used to replay prerecorded facial skin motion or behaviors [19, 35].

Z. Deng and U. Neumann, *Data-Driven 3D Facial Animation.*

Research on faces has not focused on eye movement, although the eyes play an essential role as a major channel of nonverbal communicative behavior. Eyes help to regulate the flow of conversation, signal the search for feedback during an interaction (gazing at the other person to see how she follows), look for information, express emotion (looking downward in case of sadness, embarrassment, or shame), or influence another person's behavior (staring at a person to show power) [18, 34].

Recently, eye movement has attracted attention among computer animation researchers. Directional gaze cues are frequently present to communicate the nature of the interpersonal relationship in face-to-face interactions [1]. It is estimated that 60% of conversation involves gaze and 30% involves mutual gaze [34]. Some researchers [15, 44] analyze frequencies of mutual gaze to simulate patterns of eye gaze for the participants. Social gaze serves to regulate conversation flow. Cassell and colleagues [11–13] in particular have explored eye engagement during social interactions or discourse between virtual agents. They discuss limited rules of eye engagement between animated participants in conversation. Eye movements are linked to visual attention processing: task actions generate the appropriate attentional (eye gaze or looking) behavior for virtual characters existing or performing tasks in a changing environment, such as "walk to the lamp post," "monitor the traffic light," or "reach for the box" [14].

Eye-gaze patterns for an avatar interacting with other real or virtual participants have also become important areas of study and simulation. Gaze patterns are investigated to see how observers react to whether an avatar is looking at or looking away from them [15]. Simulations for face-to-face conversation are mainly dyadic, and turn allocation using gaze signals is relatively simple. Multiparty turn-taking behavior has been less explored, and some attempts [39, 41] have been based largely on the dyadic situation. Much of this work focuses on user-perceptual issues or has involved mediated communication rather than conversational agent simulation. Intuitively, a significant difference exists in gaze behaviors between dyadic and multiparty situations: at the minimum, the latter must include mechanisms for multiple audience turn requests, acknowledgement, and attention capture.

We propose a new approach for synthesizing the trajectory kinematics and statistical distribution of saccadic eye movements. First, we present an eye movement model based on both empirical studies of saccades and statistical models of eye-tracking data. Then we address the role of gaze in multiparty conversation, giving a procedure for turn allocation, turn request, and expression of conversational feedback signals. The overview of our approach is as follows. First, we analyze a sequence of eye-tracking images in order to extract the spatiotemporal trajectory of the eye. Although the eye-tracking data can be directly replayed on a face model, its primary purpose is for deriving a statistical model of the saccades that occur. The eye-tracking video is further segmented and classified into two modes, a talking mode and a listening mode, so that we can construct a saccade model for each. The models reflect the dynamic (spatiotemporal) characteristics of natural eye movement, which include saccade magnitude, direction, duration, velocity, and inter-saccadic interval. Based on the model, we synthesize an animated face with more natural-looking and believable eye movements. Communicative aspects of eye

movement are layered on top of the saccade model to give multiparty conversational signals.

This article describes our approach in detail. Section 4.2 reviews pertinent research about saccadic eye movements and the role of gaze in communication. Section 4.3 presents an overview of our system architecture. Section 4.4 introduces our statistical model based on the analysis of eye-tracking images. An eye saccade model is constructed for both talking and listening modes and adapted for "thinking" mode. Section 4.5 shows the model implemented in agents who use appropriate social signals to simulate interactive conversations. Section 4.6 describes the architecture of our eye movement synthesis system. Finally we give our conclusions and closing remarks.

4.2 Background

4.2.1 Saccades

Saccades are rapid movements of both eyes from one gaze position to another [31]. They are the only eye movement that can be readily, consciously, and voluntarily executed by human subjects. Saccades must balance the conflicting demands of speed and accuracy, in order to minimize both time spent in transit and time spent making corrective movements.

There are a few conventions used in the eye movement literature when describing saccades. The magnitude (also called the amplitude) of a saccade is the angle through which the eyeball rotates as it changes fixation from one position in the visual environment to another. Saccade direction defines the 2D axis of rotation, with 0° being to the (person's) right. This essentially describes the eye position in polar coordinates. For example, a saccade with magnitude 10° and direction 45° is equivalent to the eyeball rotating 10° in a right-upward direction. Saccade duration is the amount of time that the movement takes to execute, typically determined using a velocity threshold. The inter-saccadic interval is the amount of time that elapses between the termination of one saccade and the beginning of the next one.

The metrics (spatiotemporal characteristics) of saccades have been well studied (for a review, see [4]). A normal saccadic movement begins with an extremely high initial acceleration (as much as $30,000^\circ/\text{sec}^2$) and terminates with almost as rapid a deceleration. Peak velocities for large saccades can be $400-600^\circ/\text{sec}$. Saccades to a goal direction are accurate to within a few degrees. Saccadic reaction time is 180–220 msec on average. Minimum inter-saccadic intervals range from 50–100 msec.

The duration and velocity of a saccade are functions of its magnitude. For saccades between 5° and 50°, the duration has a nearly constant rate of increase with magnitude and can be approximated by the linear function

$$D = D_0 + d*A, \tag{4.1}$$

where D and A are the duration and amplitude of the eye movement, respectively. The slope d represents the increment in duration per degree. It ranges from 2–2.7 msec/deg. The intercept or catch-up time D_0 typically ranges from 20–30 msec [4].

Saccadic eye movements are often accompanied by a head rotation in the same direction (gaze saccades). Large gaze shifts always include a head rotation under natural conditions; in fact, naturally occurring saccades rarely have a magnitude greater than 15° [3]. Head and eye movements are synchronous [6, 42].

4.2.2 Gaze in Social Interaction

According to psychological studies [1, 18, 26], there are three functions of gaze:

1. sending social signals: speakers use glances to emphasize words, phrases, or entire utterances while listeners use glances to signal continued attention or interest in a particular point of the speaker, or in the case of an averted gaze, lack of interest or disapproval;
2. open a channel to receive information: a speaker will look up at the listener during pauses in speech to judge how his words are being received and whether the listener wishes him to continue while the listener continually monitors the facial expressions and direction of gaze of the speaker;
3. regulate the flow of conversation: the speaker stops talking and looks at the listener, indicating that the speaker is finished and conversational participants can look at a listener to suggest that the listener be the next to speak.

Gaze aversion can signal that a person is thinking. For example, someone might look away when asked a question as she composes her response. Gaze is lowered during discussion of cognitively difficult topics. Gaze aversion is also more common while speaking as opposed to listening, especially at the beginning of utterances and when speech is hesitant. Kendon found additional changes in gaze direction, such as the speaker looking away from the listener at the beginning of an utterance and toward the listener at the end [26]. He also compared gaze during two kinds of speech pauses: phrase boundaries (the pause between two grammatical phrases of speech), and hesitation pauses (delays that occur when the speaker is unsure of what to say next). The level of gaze rises at the beginning of a phrase boundary pause, similarly to what occurs at the end of an utterance in order to collect feedback from the listener. Gaze level falls at a hesitation pause, which requires more thinking.

4.3 Overview of Eye Movement System Architecture

Figure 4.1 depicts the overall eye movement system architecture and animation procedure. First, the eye-tracking images are analyzed and a statistical eye movement model is generated using MATLAB® (The MathWorks, Inc.) (Block 1). For

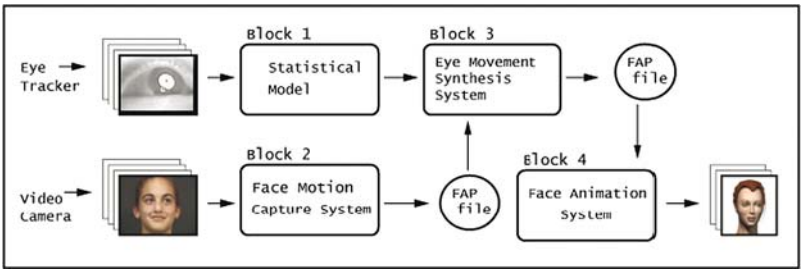


Fig. 4.1 Overall eye movement system architecture.

lip movements, eye blinks, and head rotations, we use the alterEGO face motion analysis system (Block 2), which was developed at face2face.com. The alterEGO system analyzes a series of images from a consumer digital video camera and generates a MPEG-4 Face Animation Parameter (FAP) file [24, 35]. The FAP file contains the parameter values of lip movements, eye blinks, and head rotation [35]. These components are executed offline, before the animation is created. Our principal contribution, the Eye Movement Synthesis System (EMSS) (Block 3), takes the FAP file from the alterEGO system and adds values for eye movement parameters based on the statistical model. EMSS outputs a new FAP file that contains eyeball movement as well as the lip and head movement information. We constructed the Facial Animation System (Block 4) by adding eyeball movement capability to face2face's Animator plug-in for 3D Studio Max® (Autodesk, Inc.). In other applications, such as the multiparty conversation ahead, we can output the FAP file to a different animated face model, such as the Greta head [33]. In the next section, we will explain the analysis of the eye-tracking images and the building of the statistical eye model (Block 1). More detail about the EMSS (Block 3) will be presented in Section 4.5.

4.4 Analysis of Eye-Tracking Data

4.4.1 Images from the Eye Tracker

We analyzed sequences of eye-tracking images in order to extract the spatiotemporal characteristics of the eye movements. Eye movements were recorded using a lightweight eye-tracking visor (ISCAN, Inc.). The visor is worn like a baseball cap and consists of a monocle and two miniature cameras. One camera views the visual environment from the perspective of the participant's left eye and the other views a close-up image of the left eye. Only the eye image was recorded to a digital videotape using a DSR-30 digital VCR (Sony Inc.). The ISCAN eye-tracking device measures the eye movement by comparing the corneal reflection of the light source (typically infrared) relative to the location of the pupil center. The position of the

(a) **(b)**

Fig. 4.2 **(a)** Original eye image from the eye tracker (left); **(b)** output of the Canny enhancer (right) distribution.

pupil center changes during rotation of the eye, while the corneal reflection acts as a static reference point.

The sample video we used is 9 minutes long and contains an informal conversation between two people. The speaker had used the eye-tracking device many times prior to this sample session; hence, it was not disruptive to her behaviors. The speaker was allowed to move her head freely while the video was taken. It was recorded at the rate of 30 fps. From the video clip, each image was extracted using Adobe Premiere® (Adobe Inc.). Figure 4.2(a) is an example frame showing two crosses, one for the pupil center and one for the corneal reflection.

We obtained the image (x, y) coordinates of the pupil center by using a pattern matching method. First, the features of each image are extracted by using the Canny operator [10] with the default threshold gray level. Figure 4.2(b) is a strength image output by the Canny enhancer. Second, to determine a pupil center, the position histograms along the x- and y-axes are calculated. Then the coordinates of the two center points with maximum correlation values are chosen. Finally, the sequences of (x, y) coordinates are smoothed by a median filter.

4.4.2 Saccade Statistics

Figure 4.3(a) shows the distributions of the eye position in image coordinates. The red circle is the primary position (PP), where the speaker's eye is fixated upon the listener. Figure 4.3(b) is the same distribution plotted in 3D, with the z-axis representing the frequency of occurrence at that position. The peak in the 3D plot corresponds to the primary position.

The saccade magnitude is the rotation angle between its starting position $S = (x_s, y_s)$ and ending position $E = (x_e, y_e)$, computed by

$$\theta \approx \arctan(d/r) = \arctan\left(\frac{\sqrt{(x_e - x_s)^2 + (y_e - y_s)^2}}{r}\right), \tag{4.2}$$

where d is the Euclidean distance traversed by the pupil center and r is the radius of the eyeball. The radius r is assumed to be one half of $x_{\max}$, the width of the eye-tracker image (640 pixels).

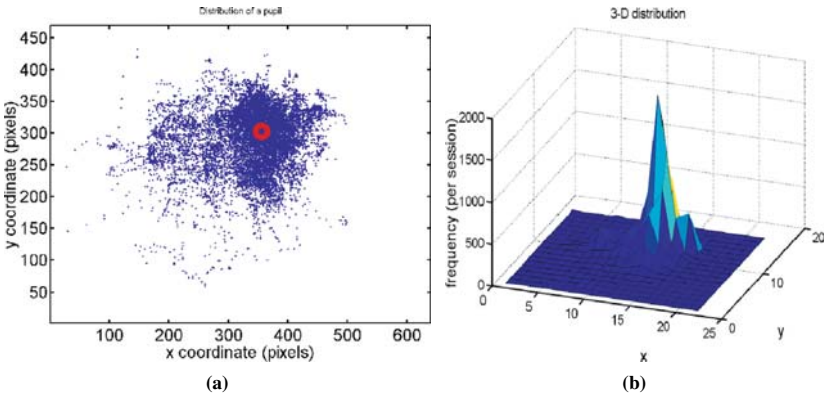


Fig. 4.3 **(a)** Distribution of pupil centers; **(b)** 3D view of same distribution.

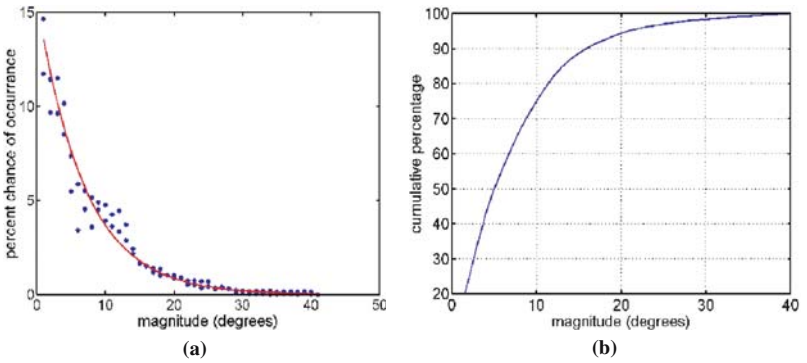


Fig. 4.4 **(a)** Frequency of occurrence of saccade magnitudes; **(b)** cumulative percentage of magnitudes

The frequency of occurrence of a given saccade magnitude during the entire recording session is shown in Fig. 4.4(a). Using a least-mean-squares criterion, the distribution was fitted to the exponential function

$$P = 15.7e^{-\frac{A}{6.9}}, \tag{4.3}$$

where P is the percent chance to occur and A is the saccade magnitude in degrees. The fitted function is used for choosing a saccade magnitude during synthesis. Figure 4.4(b) shows the cumulative percentage of saccade magnitudes: the probability that a given saccade will be smaller than magnitude x. Note that 90% of the time the saccade angles are less than 15°, which is consistent with a previous study [3].

Saccade directions are also obtained from the video. For simplicity, the directions are quantized into 8 evenly spaced bins with centers 45° apart. The distribution of saccade directions is shown in Table 4.1. One interesting observation is that up-down and left-right movements occurred more than twice as often as diagonal movements. Also, up-down movements happened equally as often as left-right movements.

Table 4.1 Distribution of saccade directions.

Direction	0°	45°	90°	135°	180°	225°	270°	315°
%	15.54	6.46	17.69	7.44	16.80	7.89	20.38	7.79

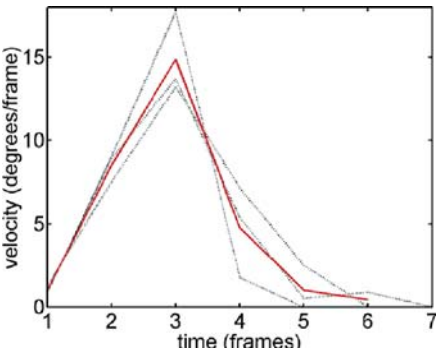


Fig. 4.5 Instantaneous velocity functions of saccades.

Saccade duration was measured using a velocity threshold of 40°/sec (1.33°/frame). The durations were then used to derive an instantaneous velocity curve for every saccade in the eye-track record. Sample curves are shown in Fig. 4.5 (black dotted lines). The duration of each eye movement is normalized to six frames. The normalized curves are used to fit a 6-dimensional polynomial (red solid line):

$$Y = 0.1251X^6 - 3.1619X^5 + 31.5032X^4 - 155.8713X^3 + 394.0271X^2 - 465.9513X + 200.3621, \tag{4.4}$$

where x is frame 1 to 6 and y is instantaneous velocity (°/frame).

The inter-saccadic interval is incorporated by defining two classes of gaze, *mutual* and *away*. In mutual gaze, the subject's eye is in the primary position, while in gaze away it is not. The duration that the subject remains in one of these two gaze states is analogous to the inter-saccadic interval. Figures 4.6(a) and (b) plot

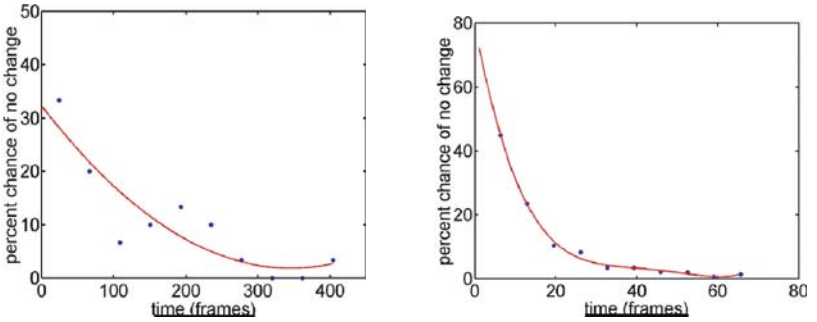


Fig. 4.6 **(a)** Frequency of mutual gaze duration while talking; **(b)** frequency of gaze away duration while talking.

duration distributions for the two types of gaze while the subject was talking. They show the percent chance of remaining in a particular gaze mode (i.e., not making a saccade) as a function of elapsed time. The polynomial fitting function for mutual gaze duration is

$$Y = 2.5524e - 4X^2 - 0.1763X + 32.2815 \tag{4.5}$$

and for gaze away duration is

$$Y = 1.8798e - 5X^4 + 0.0034X^3 + 0.2262X^2 + 6.7021X + 78.831. \tag{4.6}$$

Note that the inter-saccadic interval tends to be much shorter when the eyes are not in the primary position.

4.4.3 Talking Mode vs. Listening Mode

Characteristics of gaze differ depending on whether a subject is talking, listening, or thinking [1]. We manually segmented the video eye movement data to obtain the statistical properties of saccades in these modes. Figures 4.7(a) and (b) show the eye position distributions for talking and listening, respectively. While talking, 92% of the time the saccade magnitude is 25° or less. While listening, over 98% of the time the magnitude is less than 25°. The average magnitude is 15.64° ± 11.86° (mean±stdev) for talking and 13.83°±8.88° for listening. In general, the magnitude distribution of listening is much narrower than that of talking: when the subject is speaking, eye movements are more dynamic and active. This is also apparent while watching the eye-tracking video.

Inter-saccadic intervals also differ between talking and listening modes. While talking, the average mutual gaze and gaze away durations are 93.9 ± 94.9 frames and 27.8 ± 24.0 frames, respectively. The complete distributions are shown in Figs. 4.7(a) and (b). While listening, the average durations are 237.5 ± 47.1 frames

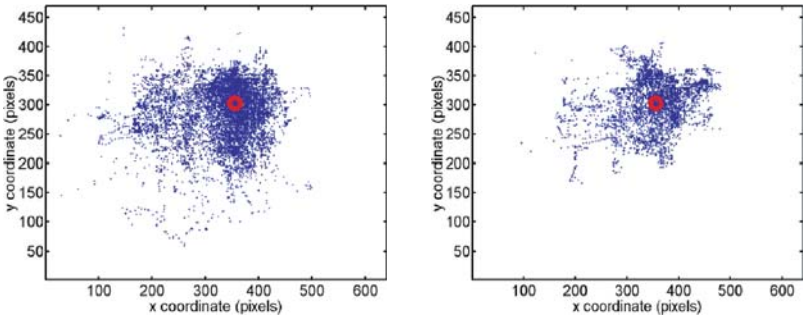


Fig. 4.7 Distribution of saccades **(a)** in talking mode; **(b)** in listening mode.

Table 4.2 Neurolinguistic information processing and corresponding eye movement patterns.

Eye Movement	Information Processing
Eye up and to the right	Trying to envision an event that has never been seen
Eye up and to the left	Recalling an event that has been seen
Unfocused eyes looking fixedly into space	Visualizing an event, real or imagined
Eye down and to the right	Sorting out sensations of the body
Eye down and to the left	Carrying on an internal conversation

for mutual gaze and 13.0 ± 7.1 frames for gaze away. These distributions were far more symmetric and could be suitably described with Gaussians. The longer mutual gaze times for listening are consistent with earlier empirical results [1] in which the speaker was looking at the listener 41% of the time, while the listener was looking at the speaker 75% of the time.

While watching a different video of a subject performing a monologue ("tell us about yourself"), we observed eye movements during periods where the subject was not speaking (and she clearly wasn't "listening" to someone else). During such subjective "thinking" modes, we found that people tend to make more eye movements upward or downward in order to avoid outside information and concentrate on their inner thoughts and emotional state. In fact, neurolinguistic programming theory postulates that the direction of eye movement is a reflection of cognitive activity [27]. This theory associates eye positions with different types of information processing (Table 4.2). Although neurolinguistic ideas often fail to survive rigorous experimental testing, the patterns for eye movement have received independent validation [9]. Remembering a has-been-seen event is significantly suggestive of a state of mind so that turning eyes up and to the left most frequently occurs when people are thinking. At that time, we observe the eyeball will have a long hold when it reaches the maximum magnitude of the current saccade. When we animate a character using the talking, listening, and thinking modes, we monitor long pauses in a speech signal as a trigger for the thinking mode and adjust the upward and downward direction distribution from the preliminary study.

4.5 Gaze Role in Multiparty Turn-Taking

Directional gaze behaviors and visual contact signal and monitor the initiation, maintenance, and termination of communicative messages [13]. Two participants use mutual gaze to look at each other, usually in the face region. Gaze contact means they look in each other's eyes. In gaze aversion, one participant looks away when others are looking toward her. Short mutual gaze ($\sim$1 sec) is a powerful mechanism that induces arousal in the other participants [27]. Gaze diminishes when disavowing social contact. By avoiding eye gaze in an apparently natural way, an audience expresses an unwillingness to speak.

Table 4.3 Turn-taking and associated gaze behaviors.

State	Signals	Gaze Behavior
Speaker	Turn yielding	Look toward listener
	Turn claiming suppression signal	Avert gaze contact from audience
	Within turn signal	Look toward audience
	No turn signal	Look away
Audiences	Back channel signal	Look toward speaker
	Turn claiming signal	Seek gaze contact from speaker
	Turn suppression signal	Avert gaze contact from speaker
	Turn claiming suppression signal	Look toward other aspiring audiences to prevent their speaking
	No response	Random

Conversation proceeds in turns. Two mutually exclusive states are posited for each participant: the speaker who claims the speaking turn and the audience who does not. Gaze provides turn-taking signals to regulate the flow of communication. Table 4.3 shows how gaze behaviors act to maintain and regulate multiparty conversations. Figure 4.8 shows sample images of the face2face.com animated face with eye movements.

In dyadic conversation, at the completion of an utterance or thought unit, the speaker gives a lengthy glance to the audience to yield a speaking turn. This gaze cue persists until the audience assumes the speaking role (Fig. 4.9(a)). The multiparty case requires a turn-allocation strategy. Inspired by Sacks [38], we address

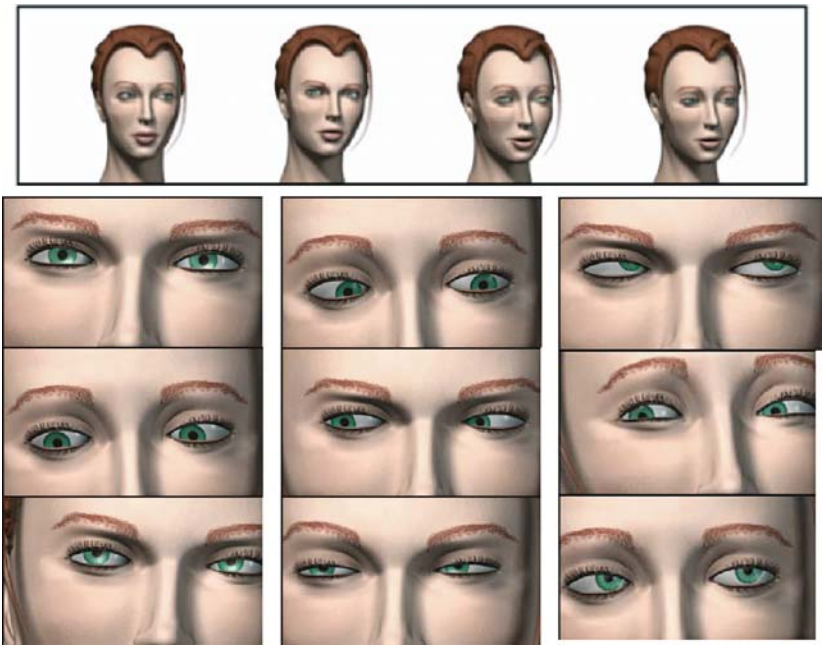
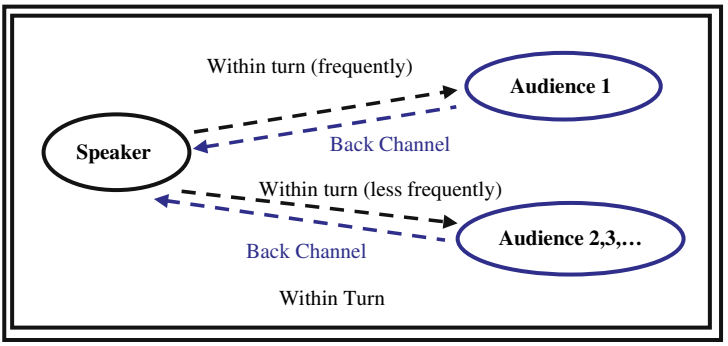
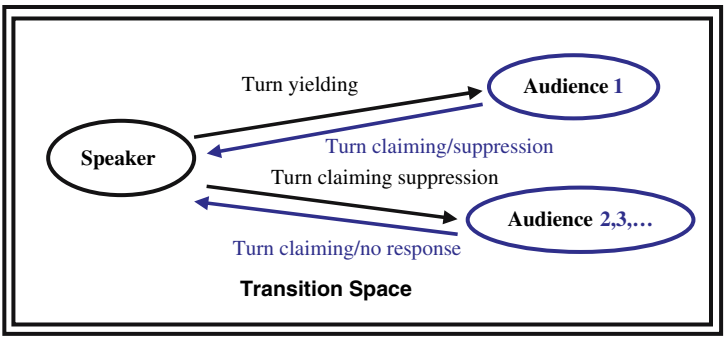


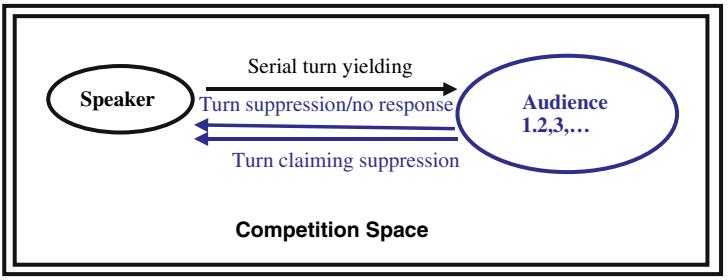
Fig. 4.8 Sample images of the face2face.com animated face with eye movements.



(1) Gaze signals within turn



(2) Gaze signals in Transition Space



(3) Gaze signals in Competition Space

Fig. 4.9 Diagrams for turn taking allocation employed conversational gaze signal.

the multiparty issue with two mechanisms: a *transition space*, where the speaker selects the next speaker, and a *competition space*, where the next turn is allocated by self-selection.

Transition space (Fig. 4.9(b))

Speaker:

1: She gives a lengthy glance (turn yielding) to one of the audiences.

2.i: Receiving gaze contact (turn claiming) from the audience, the speaker relinquishes the floor.

2.ii: Receiving gaze aversion (turn suppression) from the audience, the speaker decides to keep transition space to find another audience or go to competition space directly. If no one wants to speak, the speaker has the option of continuing or halting.

Audience:

1: An audience who wants a turn will look toward speaker's eyes to signal her desire to speak (turn claiming), and want to draw the attention of the speaker.
2: An audience receiving speaker gaze (turn yielding) uses quick gaze contact (turn claiming) to accept the turn or lengthy gaze aversion (turn suppression) to reject it.

Competition space (Fig. 4.9(c))
Speaker:
She scans all the audiences, serially sending a turn yielding signal (see Figs. 4.10(a) and (b)).

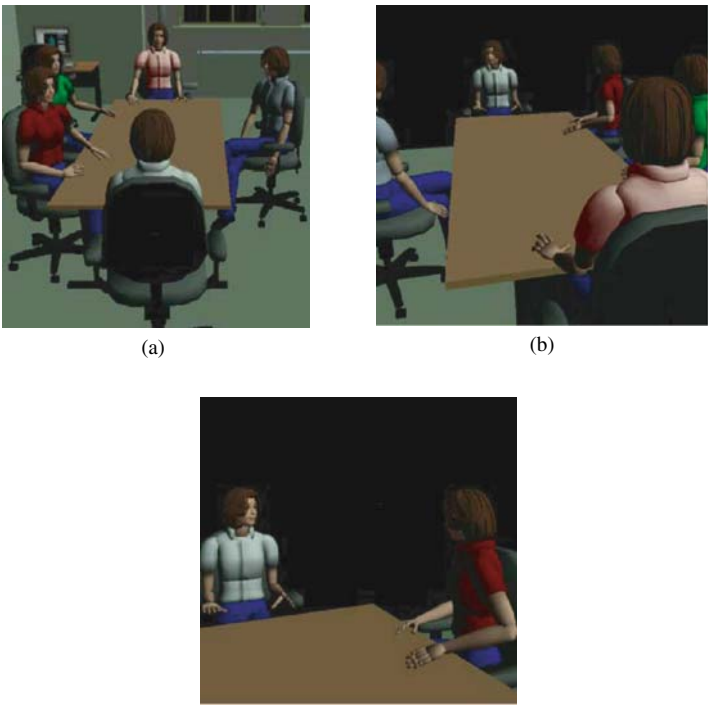


Fig. 4.10 Sample images from a five-party conversation demonstration. (a) A full-view image of five conversational agents sitting around a table; the main speaker is in the foreground with her back to the camera. (b) The main speaker sends a turn-yielding gaze signal to the agent sitting to her right. (c) The main speaker sends a turn gaze-yielding signal to the agent sitting on the first place to her left.

Audience:

They may have eye interactions at that time. The aspiring audience looks toward the speaker to signal a desire to speak (turn claiming). After receiving visual contact from the speaker, she looks at all the other aspiring audiences to signal her taking the floor (turn claiming suppression). Non-aspiring audiences may follow the speaker's gaze direction or use random gaze (no response).

Turns begin and end smoothly, with short lapses of time in between. Occasionally an audience's turn-claim in the absence of a speaker's turn signal results in simultaneous turns [27] between audiences, even between audience and speaker. Favorable simultaneous turns will occur that show it is a comfortable and communicative circumstance. The general rule is that the first speaker continues and the others drop out. The dropouts lower gaze or avert gaze to signal giving up.

Within a turn, audiences spend more time looking toward the speaker (back channel) to signal attention and interest. They focus on the speaker's face area around the eyes. The speaker generally looks less often at audiences except to monitor their acceptance and understanding (within turn signal). The speaker glances during grammatical breaks, at the end of a thought unit or idea, and at the end of the utterance to obtain feedback. The speaker usually assigns a longer and more frequent glance to the audience to whom she would like pass the floor.

4.6 Synthesis of Natural Eye Movement

A detailed block diagram of the eye movement synthesis model is illustrated in Fig. 4.11. The key components of the model consist of the (1) **Attention Monitor (AttMon)**, (2) **Parameter Generator (ParGen)**, and (3) **Saccade Synthesizer (SacSyn)**.

AttMon monitors the system state and other necessary information, such as whether it is in talking, listening, or thinking mode, whether the direction of the

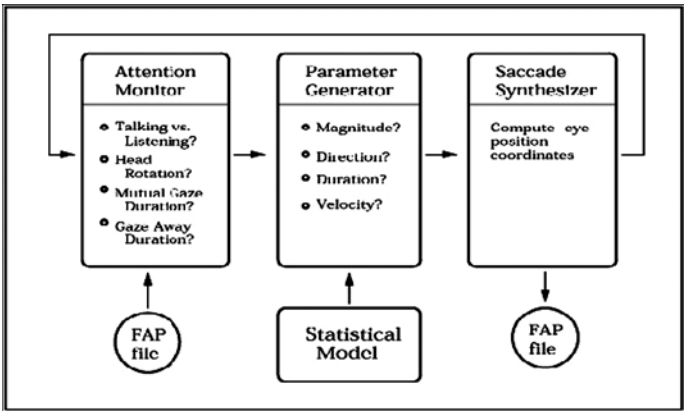


Fig. 4.11 Block diagram of the statistical eye movement model.

head rotation has changed, or whether the current frame has reached the mutual gaze duration or gaze away duration. By default, the synthesis state starts from the mutual gaze state.

If the direction of head rotation has changed and its amplitude is larger than an empirically chosen threshold, then it invokes **ParGen** to initiate eye movement. Also, if the timer for either mutual gaze or gaze away duration is expired, it invokes **ParGen. ParGen** determines saccade magnitude, direction, duration, and instantaneous velocity. It also decides the gaze away duration or mutual gaze duration depending on the current state. Then it invokes **SacSyn**, where appropriate saccade movement is synthesized and coded into FAP values.

Saccade magnitude is determined using the inverse of the exponential fitting function of Fig. 4.4(a). First, a random number between 0 and 15 is generated. The random number corresponds to the *y*-axis (percentage of frequency) in Fig. 4.4(a). The magnitude is computed from the inverse of Eq. 4.3,

$$A = -6.9 * \log(P/15.7), \tag{4.7}$$

where A is the saccade magnitude in degrees and P is the random number generated, i.e., the percentage of occurrence. This inverse mapping using a random number guarantees that the saccade magnitude has the same probability distribution as shown in Fig. 4.4(a). Based on the analysis result in Section 4.4.3, the maximum saccade magnitude is limited to 27.5° for talking mode and 22.7° for listening mode. The maximum magnitude thresholds are determined by the average magnitude plus one standard deviation for each mode.

Saccade direction is determined by two criteria. If the head rotation is larger than a threshold, the saccade direction follows the head rotation. Otherwise, the direction is determined based on the distribution shown in Table 4.1. A uniformly distributed random number between 0 and 100 is generated and 8 non-uniform intervals are assigned to the respective directions. That is, a random number between 0–15.54 is assigned to the direction 0° (right), a number between 15.54–22.00 to the direction 45° (up-right), and so on. Thus, 15.54% of the time a pure rightward saccade will occur, and 6.46% of the time an up-rightward saccade will be generated.

Given a saccade magnitude A, the duration is calculated using Eq. 4.1 with values $d = 2.4\,\text{msec/deg}$ and $D_0 = 25\,\text{msec}$. The velocity of the saccade is then determined using the fitted instantaneous velocity curve (Eq. 4.4). Given the saccade duration D in frames, the instantaneous velocity model is resampled at D times the original sample rate (1/6). The resulting velocity follows the shape of the fitted curve with the desired duration D.

In talking mode, the mutual gaze duration and gaze away duration are determined similarly to the other parameters, using inverses of the polynomial fitting functions (Eqs. 4.5 and 4.6). Using the random numbers generated for the percentage range, corresponding durations are calculated by root-solving the fitting functions. The resulting durations have the same probability distributions. In listening mode, inter-saccadic intervals are obtained using Gaussian random numbers with the duration values given in Section 4.4.3: 237.5 ± 47.1 frames for mutual gaze and 13.0 ± 7.1 frames for gaze away.

SacSyn collects all synthesis parameters obtained above and calculates the sequence of coordinates for the eye centers. The coordinate values for eye movements are then translated into FAP values for the MPEG-4 standard [24]. For facial animation, we merge the eye movement FAP values with the parameters for lip movement, head movement, and eye blinks provided by the alterEGO system. After synthesizing a saccade movement, **SacSyn** sets the synthesis state to either gaze away state or mutual gaze state. Again, **AttMon** checks the head movement, internal mode of the agent, and the timer for gaze away duration.

When a new eye movement has to be synthesized, **ParGen** is invoked to determine the next target position, e.g., another agent's face. Depending on the next target position, the state either stays at the gaze away state or returns to the mutual gaze state. In addition to applying the saccade data from the FAP file, we incorporate the vestibulo-ocular reflex (VOR). The VOR stabilizes gaze during head movements (as long as they are not gaze saccades) by causing the eyes to counter-roll in the opposite direction [31].

4.7 Conclusions

We presented eye saccade models based on the statistical analysis of an eye-tracking video. The eye-tracking video is segmented and classified into talking, listening, and thinking modes. A saccade model is constructed for each of the three modes. The models reflect the dynamic characteristics of natural eye movement, which include saccade magnitude, duration, velocity, and inter-saccadic interval. In a sample experiment with 12 observers, 10 of 12 judged the model visually and psychologically superior to two alternate methods of automatic gaze generation: no saccades and randomized saccades [28]. This model is implemented on conversational agents during face-to-face interaction. Simultaneously, the role of gaze on the turn-taking allocation strategy, appearance of awareness, and expression of the feedback signal are addressed in the simulation.

One way to generate eye movements on a face model is to replay the eye-tracking data previously recorded from a subject. Preliminary tests using this method indicated that the replayed eye movements looked natural by themselves, but were often not synchronized with speech or head movements. An additional drawback to this method is that it requires new data to be collected every time a novel eye-track record is desired. Once the distributions for the statistical model are derived, any number of unique eye movement sequences can be animated.

The eye movement video used to construct the saccade statistics was limited to a frame rate of 30 Hz, which can lead to aliasing. In practice, this is not a significant problem, best illustrated by an example. Consider a small saccade of 2°, which will have a duration of around 30 msec (Eq. 4.1). To completely lose all information on the dynamics of this saccade, it must begin within 3 msec of the first frame capture, so that it is completely finished by the second frame capture 33 msec later. This can be expected to happen around 10% of the time (3/33). From Fig. 4.5(b), it can be seen that saccades this small comprise about 20% of all saccades in the record, so only around 2% of all saccades should be severely aliased. This small percentage

has little effect on the instantaneous velocity function of Fig. 4.6. Since saccade starting and ending positions are still recoverable from the video, the magnitude and direction are much less susceptible to aliasing problems.

A more important consideration is the handling of the VOR during the eye movement recording. A change in eye position that is due to a saccade (e.g., up and to the left) must be distinguishable from a change that is due to head rotation (e.g., down and to the right). One solution is to include a sensor that monitors head position. When head position is added to eye position, the resultant gaze position is without the effects of the VOR. However, this introduces the new problem that eye and head movements are no longer independent. An alternate approach is to differentiate the eye position data, and threshold the resultant eye velocity (e.g., at 80°/sec) to screen out non-saccadic movements. Although this can be performed post-hoc, it is not robust at low sampling rates. For example, revisiting the above example, a 2° position change that occurred between two frames may have taken 33 msec (velocity = 60°/sec) or 3 msec (velocity = 670°/sec). In this study, head movements in subjects occurred infrequently enough that they were unlikely to severely contaminate the saccade data. However, in future work they must be better controlled, using improved equipment, more elaborate analysis routines, or a combination of both.

A number of enhancements to our system could be implemented in the future. During the analysis of eye-tracking images, we noticed a high correlation between the eyes and the eyelid movement that could be incorporated; Deng's model can be applied to improve this aspect of the simulation [17]. A scan-path model could be added, using not only the tracking of close-up eye images but also the visual environment images taken from the perspective of the participant's eye. Additional subjects could be added to the pool of saccade data, reducing the likelihood of idiosyncrasies in the statistical model. Other modeling procedures themselves could be investigated, such as neural networks or Markov models. Improvements such as these will further increase the realism of a conversational agent.

Acknowledgment This article is derived and extended from the paper [28] originally published by ACM. The original document may be found at http://doi.acm.org/ 10.1145/566570.566629. Permission to use material for it for this publication is greatly appreciated. Thanks to Eric Petajan, Doug DeCarlo, Minkyu Lee, Ed Roney, Jan Allbeck, Karen Carter, and Koji Ashida for their help and comments, face2face.com for the face-tracking software, John Trueswell for the eye-tracking data, and Andrew Weidenhammer for the face model and subject data. Catherine Pelachaud kindly supplied the Greta head we attached to the UGS Jack bodies running in the Lockheed-Martin Moorestown Human Testbed software. This research was partially supported by the Office of Naval Research K-5-55043/3916-1552793 and N000140410259, NSF IIS-9900297 and IIS-0200983, and NSF-STC Cooperative Agreement number SBR-89-20230.

References

1. Argyle M, Cook M (1976) *Gaze and Mutual Gaze*. Cambridge University Press, London.
2. Argyle M, Dean J (1965) Eye-contact, distance and affiliation. *Sociometry*, 28: 289–304.
3. Bahill AT, Adler D, Stark L (1975) Most naturally occurring human saccades have magnitudes of 15 degrees or less. In: *Investigative Ophthalmol. Vis. Sci*., 14: 468–469.

4. Becker W (1989) Metrics. In: RH Wurtz and ME Goldberg (eds). *The Neurobiology of Saccadic Eye Movements*, Elsevier Science Publishers BV (Biomedical Division), New York, NY, ch. 2 pp, 13–67.
5. Beeler GW (1965) Stochastic processes in the human eye movement control system. Ph.D. thesis, California Institute of Technology.
6. Bizzi E, Kalil RE, Morasso P, Tagliasco V (1972) Central programming and peripheral feedback during eye-head coordination in monkeys. *Bibl. Ophthal*., 82: 220–232.
7. Blanz V, Vetter T (1999) A morphable model for the synthesis of 3D faces. In: *Computer Graphics (SIGGRAPH Proc.)*, pp 187–194.
8. Brand M (1999) Voice puppetry. In: *Computer Graphics (SIGGRAPH Proc.)*, pp 21–28.
9. Buckner W, Reese E, Reese R (1987) Eye movement as an indicator of sensory components in thought. *J. Counseling Psych*, 34(3): 283–287.
10. Canny J (1986) A computational approach to edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(6): 679–698.
11. Cassell J, Pelachaud C, Badler N, Steedman M, Achorn B, Becket T, Douville B, Prevost S, Stone M (1994) Animated conversation: Rule-based generation of facial expression gesture and spoken intonation for multiple conversational agents. In: *Computer Graphics (SIGGRAPH Proc.)*, pp 413–420.
12. Cassell J, Torres O, Prevost S (1999) Turn taking vs. discourse structure: How best to model multimodal conversation. In: Y Wilks (ed) *Machine Conversations*. Kluwer: The Hague, pp 143–154.
13. Cassell J, Vilhjalmsson H, Bickmore T (2001) BEAT: The Behavior Expression Animation Toolkit. In: *Computer Graphics (SIGGRAPH Proc.)*, pp 477–486.
14. Chopra-Khullar S, Badler N (2001) Where to look? Automating visual attending behaviors of virtual human characters. *Autonomous Agents and Multi-Agent Systems*, 4(1/2): 9–23.
15. Colburn A, Cohen MF, Drucker SM (2000) The role of eye gaze in avatar mediated conversational interfaces. Microsoft Tech Report 2000–81.
16. DeCarlo D, Metaxas D, Stone M (1998) An anthropometric face model using variational techniques. In: *Computer Graphics (SIGGRAPH Proc.)*, pp 67–74.
17. Deng Z, Lewis JP, Neumann U (2005) Automated eye motion using texture synthesis. In: *IEEE Computer Graphics and Applications*, 25(2): 24–30.
18. Duncan S (1974) Some signals and rules for taking speaking turns in conversations. In: Weitz (ed) *Nonverbal Communication*. Oxford University Press, New York.
19. Essa I, Pentland A (1995) Facial expression recognition using a dynamic model and motion energy. In: *Proc. ICCV*, pp 360–367.
20. Faigin G (1990) *The Artist's Complete Guide to Facial Expression*. Watson-Guptill Publications, New York.
21. Guenter B, Grimm C, Wood D (1998) Making faces. In: *Computer Graphics (SIGGRAPH Proc.)*, pp 55–66.
22. Gu E (2006) Multiple influences on gaze and attention behavior for embodied agents. Ph.D. thesis, University of Pennsylvania.
23. Gu E, Badler NI (2006) Visual attention and eye gaze during multiparty conversations with distractions. In: *Proc. Intelligent Virtual Agents*, LNAI 4133, pp 193–204.
24. ISO/IEC JTC 1/SC 29/WG11 (1999) N3055/N3056. MPEG-4 Manuals.
25. Kalra P, Mangili A, Magnenat-Thalmann N, Thalmann D (1992) Simulation of muscle actions using rational free form deformations. In: *Proc. Eurographics, Computer Graphics Forum*, 2(3): 59–69.
26. Kendon A (1967) Some functions of gaze direction in social interaction. *Acta Psychologica*, 26: 22–63.
27. Knapp ML, Hall JA (1997) The effects of eye behavior on human communication. In: *Nonverbal Communication in Human Interaction*, 4th ed. Harcourt Brace, Fort Worth, TX.
28. Lee SP, Badler J, Badler N (2002) Eyes alive. *ACM Trans. on Graphics (SIGGRAPH Proc.)*, 21(3): 637–644.
29. Lee WS, Magnenat-Thalmann N (2000) Fast head modeling for animation. In: *Image and Vision Computing*, 18(4): 355–364.

30. Lee Y, Terzopoulos D, Waters K (1995) Realistic modeling for facial animation. In: *Computer Graphics (SIGGRAPH Proc.)*, pp 55–62.
31. Leigh RJ, Zee DS (2006) *The Neurology of Eye Movements*, 4th ed. Oxford University Press, New York.
32. Parke F (1982) Parameterized models for facial animation. *IEEE Computer Graphics and Applications*, 2(9): 61–68.
33. Pasquariello S, Pelachaud C (2001) Greta: A simple facial animation engine. In: 6th Online World Conf. on Soft Computing in Industrial Applications, Session on Soft Computing for Intelligent 3D Agents.
34. Pelachaud C, Badler N, Steedman M (1996) Generating facial expressions for speech. *Cognitive Science*, 20(1): 1–46.
35. Petajan E (1999) Very low bitrate face animation coding in MPEG-4. In: *Encyclopedia of Telecommunications*, 17: 209–231.
36. Pighin F, Hecker J, Lischinski D, Szeliski R, Salesin DH (1998) Synthesizing realistic facial expressions from photographs. In: *Computer Graphics (SIGGRAPH Proc.)*, pp 75–84.
37. Platt S, Badler N (1981) Animating facial expressions. In: *Computer Graphics (SIGGRAPH Proc.)*, pp 245–252.
38. Sacks H, Schegloff EA, Jefferson, G (1974) A simplest systematics for the organization of turn-taking for conversation. *Language*, 50: 696–735.
39. Vertegaal R, Slagter R, van der Veer G, Nijholt A (2000b) Why conversational agents should catch the eye. In: Summary of ACM CHI 2000 Conference on Human Factors in Computing Systems.
40. Vertegaal R, Slagter R, van der Veer G, Nijholt A (2001) Eye gaze patterns in conversations; there is more to conversational agents than meets the eyes. In: ACM CHI Conference on Human Factors in Computing Systems, pp 301–308.
41. Vertegaal R, van der Veer G, Vons H (2000a) Effects of gaze on multiparty mediated communication. In: *Proc. Graphics Interface*, Morgan Kaufmann, San Francisco, pp 95–102.
42. Warabi T (1977) The reaction time of eye-head coordination in man. *Neuroscience Letters*, 6: 47–51.
43. Waters K (1987) A muscle model for animating three-dimensional facial expression. In: *Computer Graphics (SIGGRAPH Proc.)*, pp 17–24.
44. Garau M, Slater M, Bee S, Sasse M (2001) The impact of eye gaze on communication using humanoid avatars, In *Proc. CHI*, pp. 309–316.