Generating and Describing Affective Eye Behaviors

Xia MAO†, Member and Zheng LI†‡, Nonmember

SUMMARY The manner of a person’s eye movement conveys much about nonverbal information and emotional intent beyond speech. This paper describes work on expressing emotion through eye behaviors in virtual agents based on the parameters selected from the AU-Coded facial expression database and real-time eye movement data (pupil size, blink rate and saccade). A rule-based approach to generate primary (joyful, sad, angry, afraid, disgusted and surprise) and intermediate emotions (emotions that can be represented as the mixture of two primary emotions) utilized the MPEG4 FAPs (facial animation parameters) is introduced. Meanwhile, based on our research, a scripting tool, named EEMML (Emotional Eye Movement Markup Language) that enables authors to describe and generate emotional eye movement of virtual agents, is proposed.

key words: affective computing, virtual agents, emotional eye movement, markup language

1. Introduction

Virtual agents are increasingly and widely employed in virtual environments. Emotion is proven to play an important role in human intelligence. Hence, display of emotion is an important feature of virtual agents, since it raises the users’ acceptance of a synthetic character as a communication partner. Although many researchers have devoted themselves to this direction, they mainly concentrated on facial expression generation, emotional speech synthesis and emotional gesture generation. Nevertheless, eye movement is also a crucial part of human communication. To interact with human in a rich and natural way, social interfaces need to use this communicative channel effectively. Emotions can be categorized as primary emotions and intermediate emotions. Primary emotions are joy, sadness, anger, fear, disgust and surprise. Intermediate emotions were those emotions that can be represented as the mixture of two primary emotions, or as a category of one of them. The emotions considered here were the twenty-two proposed in the OCC model [1], except happy-for and fear-confirmed. In their place, disgust and surprise were added.

In this paper, we propose an approach to synthesize emotional eye movement for primary and intermediate emotions. We present an emotional eye movement generation framework based on the parameters selected from the AU-Coded facial expression database [2] and real-time eye movement data. The overview of our approach is as follows. Firstly, we obtain eye movement parameters from the Cohn-Kanade AU-Coded facial expression database and real-time eye movement data (pupil size, blink rate and saccade). Then, a rule based method is employed to generate emotional eye movement for primary and intermediate emotions in virtual agents by MPEG-4 FAP[6]. Based on the framework, we propose the emotional eye movement markup language for describing and generating emotional eye movement in virtual agents easily.

2. Related Work

Being “a window to the mind”, the eye and its behaviors are tightly coupled with human cognitive processes. Several studies have been carried out, mostly in the Psychology and Physiology fields, to find how eye movement regulates interaction between individuals. However, these works mainly concentrated on communicative or task-related eye movement, not on how eye movement reveals affective state. To date, only few works have explicitly considered virtual agents capable of revealing its emotional states through the manner of their eye movement. Among these, Badler is probably the first, stared with the main goal to make virtual agents convey emotional information by eye movement [3]. Another interesting work is done by Lance, who realized a model of emotionally expressive head and body movement during eye movement shifts based on Gaze Warping Transformation (GWT), which is a combination of temporal scaling and spatial transformation parameters that describe the manner of an emotionally expressive eye movement shift [4]. Then he proposed a model of realistic, emotionally expressive eye movement that builds upon the GWT by improving the transformation implementation, and by adding a model of eye movement drawn from the visual neuroscience literature [5]. However, the aforementioned research only focused on analysis and synthesis of primary emotions, not on the intermediate emotions. Meanwhile, the expressive force of the emotion expression through eye movement is not enough because the pupil size, blink rate and saccade are not considered. In this paper, we describe a method to enrich human-agent interaction, focusing on analysis and synthesis of primary and intermediate emotion through eye movement in virtual agents.

3. Approach

The overall framework of our research is illustrated in Fig. 1.
Firstly, we analyze the Cohn-Kanade AU-Coded facial expression database and compute statistics about the involved FAPs [6]. For pupil size, blink rate and saccade, we use eye tracker to capture and analyze real-time eye movement data, then also compute statistics about the involved FAPs. The FAPs file contains the values of cheek, nose, eyebrow, eyelid and eyeball that related to the eye movement. Then, a rule based method is employed to generate emotional gaze behaviors for primary and intermediate emotions in a virtual agent. After that, a scripting tool that can facilitate to describe and generate emotional eye movement is proposed.

3.1 Parameters Selected from Cohn-Kanade AU-Coded Facial Expression Database

The facial action coding system (FACS) is a comprehensive and widely used method of objectively describing facial activity [7]. We can adopt FACS to describe eye movement. The AU associated with eye movement are: AU1, AU2, AU4, AU5, AU7, AU9, AU41, AU43, AU44, AU45, AU61, AU62, AU63 and AU64. These AUs can be described by MPEG-4 FAPs. Table 1 shows the relationship of FAP and AU.

| FAPs              | AU       |
|-------------------|----------|
| AU1 raise_l_eyebrow(F31) + raise_r_eyebrow(F32) |
| AU2 raise_l_o_eyebrow(F35) + raise_r_o_eyebrow(F36) |
| AU4 raise_l_o_eyebrow(F35) + raise_r_o_eyebrow(F36) + raise_l_m_eyebrow(F33) + raise_r_m_eyebrow(F34) + raise_l_j_eyebrow(F51) + raise_r_j_eyebrow(F52) + squeeze_l_eyebrow(F20) + squeeze_r_eyebrow(F38) |
| AU5 close_lEyelid(F19) + close_rEyelid(F20) |
| AU7 close_lEyelid(F21) + close_rEyelid(F22) |
| AU9 raise_nose(F60) |
| AU41 close_lEyelid(F19) + close_rEyelid(F20) |
| AU43 close_lEyelid(F19) + close_rEyelid(F20) + close_lEyelid(F21) + close_rEyelid(F22) |
| AU44 yaw_lEyeball(F23) + yaw_rEyeball(F24) + pitch_lEyeball(F25) + pitch_rEyeball(F26) |
| AU45 close_lEyelid(F19) + close_rEyelid(F20) + close_lEyelid(F21) + close_rEyelid(F22) |
| AU61 yaw_lEyeball(F23) + yaw_rEyeball(F24) + pitch_lEyeball(F25) + pitch_rEyeball(F26) |
| AU62 yaw_lEyeball(F23) + yaw_rEyeball(F24) + pitch_lEyeball(F25) + pitch_rEyeball(F26) |
| AU63 yaw_lEyeball(F23) + yaw_rEyeball(F24) + pitch_lEyeball(F25) + pitch_rEyeball(F26) |
| AU64 yaw_lEyeball(F23) + yaw_rEyeball(F24) + pitch_lEyeball(F25) + pitch_rEyeball(F26) |
| Pupil Dilate_l_pupil(F29) + Dilate_r_pupil(F30) |

In order to be able to measure eye movement FAPs in Cohn-Kanade AU-Coded facial expression database, we should define a way to describe them through the movement of feature points that lie in the eye area. Quantitative description of FAPs based on particular feature points provides the means of bridging the gap between eye movement analysis and synthesis. Each emotional image sequence in Cohn-Kanade AU-Coded facial expression database begins from a neutral or nearly neutral face and ends with an emotional expression, so we compared the neutral face eye area with the emotional face eye area. Quantitative modeling of this is implemented using the feature points showed in Fig. 2. There are 13 FAPs involved, including F19, F20, F21, F22, F31, F32, F33, F34, F35, F36, F37, F38, F63, which can be represented as f(x, y) = f1(x, 1, 2, ..., 13). It consists of distances, noted as di(x, y) where x and y is feature point in neutral face eye area and emotional face eye area respectively. We analyzed the Cohn-Kanade AU-Coded facial expression database based on this qualitative modeling method and obtained several groups of the 13 FAPs values for primary emotion eye movement. Figure 3 illustrates some key frames of eye movement we generated for primary emotion.

3.2 Parameters Selected from Real-Time Eye Tracker Data

Research has demonstrated that the pupil size [8], [9], blink rate [10] and saccade [11] have correlations with psychological states. However, all the research mentioned above are from a psychological view and focused on only one aspect of eye movement and they did not concentrate on emotional.
eye movement and were not employed in emotional eye movement generation in virtual agents. In our research, we employ the an SMI IVIEW eye-tracker [12] to capture real-time eye movement from subjects to extract the spatiotemporal characteristics of the eye movements for positive and negative emotions. We can get the parameters for AU43, AU44, AU45, AU61, AU62, AU63 and AU64. Then, pupil, blink and saccade movement can be synthesized and coded into appropriate FAP values.

3.2.1 Participants

Fifty voluntary students (25 females and 25 males, mean age 23.5 years, range 18 to 30 years) were recruited from our university. All subjects had normal or corrected-to-normal vision, normal hearing and extrovert personality by their own report.

3.2.2 Stimuli

Seventy-five pictorial stimuli, 25 per category were selected from the IAPS (International Affective Picture System) [13]. There were three types of target stimuli: neutral (nonemotional), positive (which depicted scenes of affection), and negative (which depicted scenes of threat or injury). The respective means for valence for the different categories were for negative (2.5), neutral (5.0), and positive (7.4). The luminosity of the selected pictures was modified such that the luminosity values for those pictures are same. The pictorial stimuli is accompanied by the audio stimuli which were selected from the IADS (International Affective Digitized Sounds) [14]. The categories were negative highly arousing (e.g. harsh shouts), neutral (e.g. peaceful music), and positive highly arousing (e.g. bright music) stimuli. The respective means for valence for the different categories were for negative (2.4), neutral (5.2), and positive (7.1). Each positive audio is paired with one positive picture and the valence of the audio is chosen as same with the picture valence as possible (same with the negative and neutral situation). We employed the mean valence of the picture and audio to represent the pair valence. In order to facilitate the data analysis, we transferred the valence of each pair to a eleven-point (−5 to 5) scale. The lower end represented a very negative emotional stimulus, and the upper end represented a very positive emotional stimulus. The center of the scale represented a neutral stimulus. All the stimuli were about 10 s long.

3.2.3 Procedure

Firstly, subject was informed of the purpose of the study and then she/he was comfortably seated in an adjustable chair in front of a computer screen and the headphones were put on. The meaning of valence was explained as the feeling of positive or negative produced by a stimulus, and some examples were given. The eye tracker was calibrated and the subject was instructed to look at a fixation cross at the center of the screen. Twenty seconds from the onset of the fixation point the first stimulus was delivered. Ten seconds after the stimulus offset the subject heard a beep and the fixation point disappeared, which indicated that the subject had an opportunity to have a rest. Ten seconds later, the subject heard the beep again, and the fixation cross reappeared to signal the end of the resting period. Following this there was a randomized pause of 5 seconds before the next stimulus was delivered. The pupil size, blink rate and saccade were monitored by real-time eye tracking system SMI IVIEW.

3.2.4 Data Analysis

Pupil size: In the experiment, the eye tracker recorded all the subjects’ pupil sizes during the positive, negative and neutral stimuli. We got the mean pupil size value of the 50 subjects for positive, negative and neutral stimuli. It denotes that the pupil size is larger in positive and negative state than in neutral state. Every stimuli’s mean pupil size of 50 subjects are calculated. It is observed that the stronger a stimulus is rated in terms of valence, whether positive or negative, the larger the pupil size. It also can conclude from the data that pupil size is linearly related to the valence of emotion. The statistics of the pupil size are shown in Table 2. The values of smallest or biggest mean average value of smallest or biggest per subject.

For the data during the stimuli, a 2 × 3 (gender × stimulus category) ANOVA showed a significant main effect of gender $F(1, 48) = 6.3$, $p < 0.05$ and a significant main effect of stimulus category $F(2, 96) = 8.2$, $p < 0.01$ on the pupil size. The interaction of the main effects was not significant $F(2, 96) = 1.2$, $p > 0.2$. Independent pairwise comparisons showed that the pupil size was significantly larger for female than for male subjects during neutral stimuli $t = 2.9$, $df = 48$, $p < 0.05$. There were no gender differences during either positive $t = 2.2$, $df = 48$, $p > 0.1$ or negative $t = 1.2$, $df = 48$, $p > 0.5$ stimuli.

Because pupil size variation was related to gender only during the neutral stimuli we proceeded to analyze the effects of stimuli on a general level (i.e. without the gender). A one-way ANOVA for the data during the different stimuli showed a significant effect of stimulus category on the pupil size $F(2, 98) = 8.5$, $p < 0.01$. Pairwise comparisons showed that the pupil size was significantly bigger during the negative than during the neutral stimuli $t = 5.6$, $df = 49$, $p < 0.001$. The pupil size was also significantly bigger during positive than neutral stimuli $t = 5.2$, $df = 49$, $p < 0.001$. The difference between negative and positive stimuli was not significant $t = 0.3$, $df = 49$, $p > 0.5$.

Blink rate: Blink has also been related to emotional reactions. We also got the mean blink rate of the 50 subjects

| Emotion | Average | Smallest | biggest |
|--------|--------|---------|---------|
| Neutral | 4.1 ± 0.6 mm | 3.1 mm | 5.3 mm |
| Positive | 6.1 ± 0.7 mm | 4.6 mm | 7.6 mm |
| Negative | 6.2 ± 0.5 mm | 5.0 mm | 7.2 mm |

Table 2 Statistics of the pupil size.
for positive, negative and neutral stimuli. We can conclude that the blink rate drops in positive and negative state compared to neutral state. Every stimulus’s mean blink rate of 50 subjects are calculated. It shows that the blink rate decreases when the stimulus strengthens, whether positive or negative. It also can conclude from the data that blink rate is linearly related to the valence of emotion.

Table 3 shows the statistics of the blink data collected in our experiment. Column 2 reveals the average blink rate over all subjects per 10-s interval with a standard deviation. The corresponding interblink time is illustrated in column 3. Column 4 and column 5 denotes the average value of slowest and fastest blink rate per subject per 10-s respectively.

For the data during the stimuli, a $2 \times 3$ (gender x stimulus category) ANOVA showed no significant main effect of gender $F(1,48) = 1.1, p > 0.2$ and a significant main effect of stimulus category $F(2, 96) = 9.6, p < 0.01$ on the blink rate.

Because blink rate variation was not related to gender we proceeded to analyze the effects of stimuli on a general level (i.e. without the gender). A one-way ANOVA for the data during the different stimuli showed a significant effect of stimulus category on the blink rate $F(2, 98) = 9.5, p < 0.01$. Pairwise comparisons showed that the blink rate was significantly slower during the negative than during the neutral stimuli $t = 5.3, df = 49, p < 0.001$. The blink rate was also significantly slower during positive than neutral stimuli $t = 7.8, df = 49, p < 0.001$. The difference between negative and positive stimuli was not significant $t = 0.6, df = 49, p > 0.5$.

**Saccade:** Saccade is rapid movement of both eyes from one gaze position to another. Magnitude and direction are the conventions used in the eye movement literature when describing saccade.

Saccade magnitude is the angle through which the eyeball rotates as it changes fixation from one position to another. Lee made a deep research of the saccade magnitude [15]. We analyzed the saccade according to the Lee’s method and concluded that 90% of the time the saccade angles are less than 20 degree when the subjects in positive and negative emotion state, which is 90% consistent with Lee’s study.

Saccade direction defines the 2D axis of rotation, with 0 degree being to the (person’s) right. This essentially describes the eye position in polar coordinates. The observation is that diagonal movements occurred more in negative emotions than in positive ones. Also, up-down and left-right movements happened more in positive emotions than in negative ones. While in neutral state, the gaze position usually fixed to the straight direction. The distribution of saccade directions for positive and negative emotions is shown in Table 4.

For the data during the stimuli, a $2 \times 3$ (gender x stimulus category) ANOVA showed no significant main effect of gender $F(1,48) = 1.3, p > 0.5$ and a significant main effect of stimulus category $F(2, 96) = 10.6, p < 0.01$ on the saccade.

Because saccade variation was not related to gender we proceeded to analyze the effects of stimuli on a general level (i.e. without the gender). A one-way ANOVA for the data during the different stimuli showed a significant effect of stimulus category on the saccade $F(2, 98) = 125.6, p < 0.001$. Pairwise comparisons showed that the saccade variation during the negative stimuli was significantly different from during the neutral stimuli $t = 8.8, df = 49, p < 0.001$. The saccade variation in positive stimuli was also significantly different from in neutral stimuli $t = 8.2, df = 49, p < 0.001$. The difference between negative and positive stimuli was significant $t = 7.3, df = 49, p < 0.001$.

### 3.3 Synthesis Rules

After getting the 13 FAPs parameters from the AU-Coded facial expression database for primary emotion and realtime eye movement data for positive and negative emotion, a rule-based approach is employed to generate primary and intermediate emotional eye movement. This approach employs the theory brought forward by Whissell who proposed a model where each emotion has activation and evaluation values, used to locate the emotions in a coordinate system [16]. Whissell suggests that emotions are points in a space spanning a relatively small number of dimensions, which with a first approximation, seem to occupy two axes: activation and evaluation. Activation is the degree of arousal associated with the term and evaluation is the degree of pleasantness associated with the term. From the practical point of view, evaluation seems to express internal feelings of the subject and its estimation through face formation is intractable. On the other hand, activation is related to facial muscles’ movement and can be easily estimated based on facial characteristics. Therefore, we only employ each emotion’s activation value to make rules described below. Meanwhile, we utilize both the emotion wheel and the angular [17] measure to generate primary and intermediate emotional eye movement. According to the emotion wheel, emotions are distributed as a circle. We can locate each emotion’s position in the circle. The position can reflect

| Emotion | Average | Interblink time | Slowest | Fastest |
|---------|---------|----------------|---------|---------|
| Neutral | 4.4 ± 0.9 | 2.3 ± 1.2 | 2.5 | 6.7 |
| Positive | 3.3 ± 0.8 | 3.0 ± 0.7 | 1.4 | 4.9 |
| Negative | 2.8 ± 0.7 | 3.6 ± 0.5 | 1.5 | 4.5 |

| Direction | 0° | 45° | 90° | 135° | 180° | 225° | 270° | 315° |
|-----------|-----|-----|-----|------|------|------|------|------|
| Positive  | 14.58 | 8.86 | 25.79 | 9.36 | 15.67 | 6.33 | 10.67 | 8.74 |
| Negative  | 3.63  | 23.69 | 5.39 | 15.34 | 4.65  | 16.53 | 10.23 | 20.54 |
the relationship among primary and intermediate emotions. We get the Whissell’s mean values of all values of activation and evaluation, $\overline{a} = 4.5$ and $\overline{e} = 3.7$. Then, we found the primary and intermediate emotion’s angle with respect to the X-axis by:

$$\alpha = \arctan \frac{a - \overline{a}}{e - \overline{e}}$$ \hspace{1cm} (1)

For the 13 FAPs, knowing the intermediate emotion we want to obtain, we select the primary emotions located in its neighborhood (according to the angular measure), and combine them to generate its FAPs.

**Rule one:** If the angle distance of one of the primary emotions with respect to the intermediate emotion exceeds 45 degree, it is not considered. If it is the only primary emotion with respect to the intermediate emotion not exceeds 45 degree, the values of 13 FAPs of the intermediate emotion are obtained by:

$$f_{i_{\text{intermediate}}} = \frac{a_{\text{intermediate}}}{a_{\text{primary}}} f_{i_{\text{primary}}}$$ \hspace{1cm} (2)

(i = 1, 2, ..., 13), where $f_{i_{\text{intermediate}}}$ and $f_{i_{\text{primary}}}$ represents the FAPs value of intermediate emotion and primary emotion respectively. The $a_{\text{intermediate}}$ and $a_{\text{primary}}$ denote the the activation values of the intermediate and primary emotion.

**Rule two:** If the differences between the angles of both primary emotions and the angle of the intermediate emotion are less than 45 degree, the values of 13 FAPs of the intermediate emotion are obtained by:

$$f_{i_{\text{intermediate}}} = \frac{a_{\text{intermediate}}}{a_{\text{meanprimary}}} f_{i_{\text{meanprimary}}}$$ \hspace{1cm} (3)

(i = 1, 2, ..., 13), where $a_{\text{meanprimary}}$ denotes the mean activation values for both primary emotions. $f_{i_{\text{meanprimary}}}$ represents the mean values of FAPs for both primary emotions.

After getting the 13 FAPs values for the primary emotion and intermediate emotion, we will add pupil size, blink rate and saccade information to the eye movement. The rule is described as follows:

**Pupil size:**

From the Sect. 3.2, we can conclude from the data that pupil size is linearly related to the valence of stimuli. Meanwhile, the higher the valence is, the higher the activation is. Therefore, the pupil size is also linearly related to the activation of the emotion. So we propose the equation as:

$$\text{Emotion}_{pupil} = \frac{a}{\overline{a}} \overline{pupil}$$ \hspace{1cm} (4)

where $\overline{pupil}$ depicts the mean pupil size we got in the Sect. 3.2. The value is 6.1 mm and 6.2 mm for the positive and negative emotion respectively. $\overline{a}$ is the mean values of all values of activation 4.5. The $\text{Emotion}_{pupil}$ is the primary and intermediate emotions’ pupil size we want to get. $a$ is this emotion’s activation value. Note that if the value exceeds the max value or min value in Table 2, it will be set to the extremum value.

**Blink rate:**

From the Sect. 3.2, we can conclude from the data that blink rate is linearly related to the valence of stimuli. Meanwhile, the higher the valence is, the higher the activation is. Therefore, the blink rate is also linearly related to the activation of the emotion. So we propose the equation as:

$$\text{Emotion}_{blink} = \frac{a}{\overline{a}} \overline{blink}$$ \hspace{1cm} (5)

where $\overline{blink}$ denotes the mean blink rate. The value is 3.2 per-10 s and 2.8 per-10 s for the positive and negative emotion respectively as we got in the Sect. 3.2. $\overline{a}$ and $a$ are the same as in Eq. (4). The $\text{Emotion}_{blink}$ is the primary and intermediate emotions’ blink rate we want to get. Note that if the value exceeds the max value or min value in Table 3, it will set to be the extremum value.

**Saccade:**

The rule for saccade generation is a modified version of the method described by Lee [15]. The saccade magnitude can be obtained from the function of Eq. (6).

$$A = -6.9 \ast \log(P/15.7)$$ \hspace{1cm} (6)

where $A$ is the saccade magnitude in degrees and $P$ is the random number generated, i.e., the percentage of occurrence. A random number for $P$ between 0 and 20 is generated. According to the Lee’s research, 90% of the time the saccade angles are less than 15 degree during listening and talking mode. While taking the emotion into consideration, our experiment showed that 90% of the time the saccade angles are less than 20 degree during positive and negative stimuli. So the generated random number for $P$ between 0 and 20 guarantee that the saccade magnitude has the same probability distribution, namely, 90% of the time the saccade angles are less than 20 degree when the subjects in positive and negative emotional state.

The direction is determined based on the distribution shown in Table 4. A uniformly distributed random number between 0 and 100 is generated and 8 non-uniform intervals are assigned to the respective directions. That is, for negative emotion, a random number between 0–3.63 is assigned to the direction 0 degree (right), a number between 3.63–27.32 to the direction 45 degree (up-right), and so on. Thus, 3.63% of the time a pure rightward saccade will occur, and 23.69% of the time an up-rightward saccade will be generated. The direction for positive emotion state is generated employing the same way.

Figure 4 shows some key frames of the animated face with eye movement for some intermediate emotions based on our rules.

3.4 Emotional Eye Movement Markup Language

Based on the research introduced above, we propose a scripting language that enables authors to describe and generate emotional eye movement of virtual agents. The language is specifically designed for non-expert (average) users...
allowing them to direct the virtual agents’ eye movement easily. The tags defined is illustrated in Fig. 5. The tags can be classified into four categories at present: introductory tags, presentation control tags, basic eye movement tags and emotion-related tags.

The root element of an EEMML script is eemml. The root element’s tagging structure
<eemml>
contains all other tagging structures of the script. The head element specifies general information, the title, the meta (author’s information) and the agent that will perform as a presenter. The tag body refers to the sequence of events comprising the actions of virtual agents.

The seq and par elements, which belong to the presentation control tags, are responsible for events that should be executed sequentially or in parallel, respectively. They are inspired by the corresponding tags in SMIL [18].

The eyeleft, eyeright, eyetop and eyetopdown elements are employed to control the agents’ both eyes turn around, while the lookleft, lookright, lookup and lookdown elements are employed to control the agents’ both eyes and head turn around. The eyebrow element is defined to control the the agents’ eyebrow, including inner part, medial part and outer parts for both eyes. The eyelid is employed to control the agents’ eyelid, including lower part and upper parts for both eyes. The blink element is also a sub-element of the eyelid used to control the eye blink rate. The eyeball element is specified to control the agents’ eyeball, including pupil size, magnitude and direction for both eyes. With the speak element, the agent speaks one or several sentences through a TTS (text-to-speech) systems. The mark element can be used to set an arbitrary mark at a given place in the text, so that an engine can report back to the calling application that it has reached the given location.

The emotioneye element is defined to control the emotional eye movement, twenty-two kinds of emotions can be generated. All the parameters related to the twenty-two kinds emotions are obtained through the emotional eye movement generation framework described above. If the scripting author wants to modify some parameters, they can use the basic eye movement control tags below the <emotioneye> tag, then the related parameters value will be replaced. As the example shows below, the blink rate value in joy emotion will be replace by 3 per-10 s.

<emotioneye type=”joy”>
<blink rate =”3”/>
</emotioneye>

A simple example of EEMML is shown as follow. The agent will express “sad” emotion when she speaks “I’m sorry to hear that you have met an accident” then express “joy” emotion when she speaks “but you looks much better now”. The “speak-begin” and “speak-end” in tag emotion-eye represent when the emotional eye movement will begin and end respectively.

<eemml>
<head>
<title>EEMML sample</title>
<meta authorname =”David”/>
<agent id =”1” name=”Alice”/> </head>
<body>
<par>
<speak id =”1”>
<mark id =”1”>I’m sorry to hear that you caught a cold. <mark id =”2”>If I reminded you to put on the coat, that will not happen. <mark id =”3”>but you looks much better now<mark id =”4”/></speak>
</mark>
</seq>
</emotioneye type=”sad” speak-id = “1” speak-end="2”/>
</emml>
Fig. 6 Some key frames of intermediate emotion eye movement based on Raouzaiou's work.

| emotion  | Per.1 | Per.2 | emotion  | Per.1 | Per.2 |
|----------|-------|-------|----------|-------|-------|
| gratification | 34    | 32    | sadness  | 88    | 91    |
| joy      | 91    | 95    | pity     | 56    | 63    |
| pride    | 65    | 67    | admiration| 34    | 35    |
| disappoint| 67    | 71    | gratitude| 45    | 48    |
| anger    | 84    | 89    | hate     | 76    | 78    |
| disgust  | 78    | 83    | fear     | 55    | 62    |
| resentment| 78   | 81    | shame    | 76    | 85    |
| remorse  | 68    | 74    | hope     | 67    | 73    |
| surprise | 70    | 78    | reproach | 45    | 56    |
| relief   | 53    | 57    | love     | 47    | 56    |
| satisfaction | 23   | 35    | gloating| 45    | 41    |

Table 5 Percentage of recognition of primary and intermediate emotions.

4.1 Experiment 1

We evaluated the emotional eye movement generation framework. This was achieved through the subjective evaluation of primary and intermediate emotions expression in a synthetic female face “Alice”. Firstly, fifty subjects were asked to judge twenty-two videos of emotion generated by Raouzaiou’s work [19] which did not consider the eye movement when synthesized the facial expression, identifying which emotion was being demonstrated in the twenty-two kinds of emotions. Figure 6 shows some key frames of the animated face some intermediate emotions based on Raouzaiou’s work. Then, the twenty-two videos of facial expression generated by Raouzaiou’s work were modified by our framework. Figure 4 denotes some key frames of the animated face some intermediate emotions based on our work. The subjects were asked to judge the new twenty-two videos of emotion, identifying which emotion was being demonstrated in the twenty-two kinds of emotions again. The results are shown in Table 5, denoting that the emotional eye movement expressions generated by our framework acted well to enhance the expressiveness of the emotion expression and the percentage of recognition of primary and intermediate emotions has been improved more or less (except for gratification and love). Paired samples test also showed that the recognition rate was significantly larger when adding the eye movement information ($t = -5.8$, $df = 21$, $p < 0.001$). The primary emotions have higher recognition rate than intermediate emotions. Some intermediate emotions have very low recognition rate such as gratification and gloating. Intermediate emotions are those emotions that can be represented as the mixture of two primary emotions, or as a category of one of them, so gratification and gloating are always recognized as joy. In general, the results suggested that the eye movement contributed to the identification of facial expression and our framework worked well enough to bring measurable benefit to the human-agent interaction.

4.2 Experiment 2

In the following experiment, we synthesized emotional eye movements on the female face “Alice” to compare which aspect is more important in pupil size, blink rate and saccade. Firstly we synthesized the agent emotion expression with the parameters selected from the AU-Coded facial expression database as a baseline. Then, we added the eyeball information to the baseline based on our framework. Three groups of emotional eye movements were generated including twenty-two emotions. The pupil size information was added to the first group. The blink rate and saccade information was added to the second and third group respectively. The three groups were presented to thirty subjects. Then they were required to evaluate the three groups emotional eye movements against three factors: expressiveness, friendliness and intelligence. For each aspect, subjects were asked to evaluate how much they agreed with the statement, on a scale from 1 (strongly disagree) to 7 (strongly agree). The quantitative results of the questionnaire are shown in Fig. 7. The length of each bar indicates the mean score, which is also written at the right end of each bar.

For Factor 1 (expressiveness), the multiple Friedman comparison suggests that the agent with saccade information are judged more expressive ($R_{saccade} > R_{blink} > R_{pupil}$).

For Factor 2 (friendliness), the multiple Friedman comparison denotes that the agent with blink rate information are judged more friendly ($R_{blink} > R_{saccade} > R_{pupil}$).

For Factor 3 (intelligence), the multiple Friedman comparison means that the agent with saccade information are judged more intelligent ($R_{saccade} > R_{blink} > R_{pupil}$).

The factor analysis showed that in three factors (expressiveness, friendliness and intelligence) pupil size is the least important. It is mainly because pupil size is small
4.3 Experiment 3

This experiment is set up from the point of view of the EEMML. Considered that this language is specifically designed for non-expert (average) users allowing them to direct the virtual agents’ eye movement easily, thirty subjects who never employed the scripting languages were asked to learn and use EEMML to describe and generate the emotional eye movement in virtual agents. Then they were required to evaluate the language against three aspects: usage, function and adoption. For each aspect, users were asked to evaluate how much they agreed with a statement, on a scale from 1 (strongly disagree) to 5 (strongly agree). The quantitative results of the questionnaire are shown in Fig. 8. The graphic shows the percentage of the ratings attributed by the subjects.

- Q1: The language is easy of use.
- Q2: The language is effective for programming complex emotional eye movement in virtual agents.
- Q3: I am pleased with the language and would use it to script the emotional eye movement in virtual agents.

Analyzing the graphics, we perceive that majority of the subjects agreed with our statements which evaluated the more general aspects of the language more or less. For all these items, the rating given by the subjects were positive (“Normal” to “Strongly Agree” beyond at least 90 percent of the evaluations for each statement). The results of the experiment demonstrated that the EEMML system is a convenient tool to describe and generate emotional eye movement in virtual agents and has potential in the human-agent interaction.

In the post-study interview the users gave their informal comments about EEMML which suggested some important areas for improvement. The remarks below provide the suggestions most relevant to the scope of our work.

Several subjects remarked that the head motion should be described in EEMML because it played an important role during eye movement. Another subject suggested that we should consider the humidity of the eyes (dry/wet/tears). Two subjects complained that the system should provide more virtual agents to choose. A future research might consider to provide head motion and humidity control tag and more MPEG-4 FAP based virtual agents will be added to our system.

5. Discussion

Non-verbal behavior, particularly eye movement, plays a fundamental role in nonverbal communication among people. Considered that eye movement conveys much about emotional intent beyond speech, we propose a framework that enables virtual agents to express emotion through eye movement. The EEMML is intended as a convenient tool to describe and generate emotional eye movement in virtual agents. As a scripting language, EEMML is located at a high level that allows easy control of emotional eye movement sequences based on the framework aforementioned. Almost all the eye related parameters including eyebrow, eyelid and eyeball can be controlled by EEMML. The EEMML can be integrated into our MIML system [20] as a submodule. That will make the MIML system a powerful and easy-to-use virtual agents scripting tool for non-expert users to script the attractive and believable intelligent human-agent interaction.

6. Conclusion and Future Work

In this paper, we present a computational framework that enables virtual agents to express emotion through eye movement based on the parameters selected from the AU-Coded facial expression database and real-time eye movement data (pupil size, blink rate and saccade). Based on the framework, we also introduce the emotional eye movement markup language for describing and generating emotional eye movement in virtual agents easily. We believe our approach offers a different view in relation to other emotional eye animation models found in the literature. Moreover, by analyzing the users’ opinions in our evaluation procedure, we concluded that our system worked well enough to bring measurable benefit to the human-agent interaction.

There are a number of enhancements to our framework which could be implemented in the future. We plan to employ the motion capture system to get the eye related feature point parameters instead of analyzing the AU-Coded facial expression database. Although the MPEG-4 based model appears useful, it lacks representation of texture or appearance given that “eye movement” is conceptualized broadly to include most changes in eye and face movement in the mid and upper face. We intend to do some research from this point of view. Improvements such as these will further increase the realism of virtual agents.
Acknowledgements

This work is supported by the National Research Foundation for the Doctoral Program of Higher Education of China (No.2007006057) and National Nature Science Foundation of China (No.60873269).

References

[1] A. Ortony, G.L. Clore, and A. Collins, The Cognitive Structure of Emotions, Cambridge Univ. Press, 1988.
[2] J.F. Cohn, A.J. Zlochower, J. Lien, and T. Kanade, “Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding,” Psychophysiology, vol.36, pp.35–43, 1999.
[3] E. Gu and N.I. Badler, “Visual attention and eye gaze during multiparty conversations with distractions,” Proc. International Conference on Intelligent Virtual Agents, pp.193–204, 2007.
[4] B. Lance and S. Marsella, “Emotionally expressive head and body movements during gaze shifts,” Proc. International Conference on Intelligent Virtual Agents, pp.72–85, 2007.
[5] B. Lance and S. Marsella, “The relation between gaze behavior and the attribution of emotion: An empirical study,” Proc. International Conference on Intelligent Virtual Agents, pp.1–14, 2008.
[6] MPEG4: Moving Pictures Expert Group, ISO/IEC 14496, International Standard on Coding of Audio-Visual Objects.
[7] P. Ekman and W.V. Friesen, Manual for the Facial Action Coding System, 1976.
[8] T. Partala and V. Surakka, “Pupil size variation as an indication of affective processing,” Int. J. Human-Computer Studies, vol.59, pp.185–198, 2003.
[9] M.M. Bradley, L. Miccoli, M.A. Escrib, and P.J. Lang, “The pupil as a measure of emotional arousal and autonomic activation,” Psychophysiology, vol.45, pp.602–607, 2008.
[10] K. Takashima, Y. Miyata, Y. Itoh, Y. Kitamura, and F. Kishino, “Effects of avatar’s blinking animation on person impressions,” Proc. Graphics Interface, pp.169–176, 2008.
[11] J. Kissler and A. Keil, “Look-don’t look! How emotional pictures affect pro- and anti-saccades,” Exp Brain Res, vol.188, pp.215–222, 2008.
[12] SMI: http://www.smivision.com
[13] P.J. Lang, M.M. Bradley, and B.N. Cuthbert, “Affective ratings of pictures and instruction manual,” Technical Report A-8. University of Florida, Gainesville, FL, 2005.
[14] M.M. Bradley and P.J. Lang, “International affective digitized sounds (IADS), Stimuli, instruction manual and affective ratings,” Tech. Rep. no.B-2, The Center for Research in Psychophysiology, University of Florida, Gainesville, FL, 1999.
[15] S. Lee, J. Badler, and N. Badler, “Eyes alive,” ACM Trans. Graphics, vol.21, no.3, pp.637–644, 2002.
[16] C. Whissel, “The dictionary of affect in language,” in Emotion: Theory, Research and Experience, pp.113–131, Academic Press, 1989.
[17] R. Plutchik, Emotion: A Psychoevolutionary Synthesis, Harper and Row, New York, 1980.
[18] SMIL Home Page: URL: http://www.w3.org/AudioVideo/
[19] A. Raouzaiou, N. Tsapatsoulis, K. Karpouzis, and S. Kollias, “Parameterized facial expression synthesis based on MPEG-4,” EURASIP J. Applied Signal Processing, vol.10, pp.1021–1038, 2002.
[20] X. Mao, Z. Li, and HY. Bao, “Extension of MPML with Emotion Recognition Functions,” Proc. IVA2008, pp.289–295, 2008.