Providing Expressive Eye Movement to Virtual Agents

Zheng Li
School of Electronic and Information Engineering
Beihang University
Beijing, China
buaa_david@yahoo.com.cn

Xia Mao
School of Electronic and Information Engineering
Beihang University
Beijing, China
moukyoucn@yahoo.com.cn

Lei Liu
School of Electronic and Information Engineering
Beihang University
Beijing, China
liulei032@163.com

ABSTRACT

Non-verbal behavior, particularly eye movement, plays a fundamental role in nonverbal communication among people. In order to realize natural and intuitive human-agent interaction, the virtual agents need to employ this communicative channel effectively. Against this background, our research addresses the problem of emotionally expressive eye movement manner by describing a preliminary approach based on the parameters picked from real-time eye movement data (pupil size, blink rate and saccade).

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems-Human factors; H5.2 [**Information Interfaces and Presentation**]: User Interfaces

General Terms

Human Factors

Keywords

Nonverbal behavior, virtual agent, eye movement

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 parter. 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.

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[1]. 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[2]. However, the aforementioned research only focused on analysis and synthesis of primary emotions (joy, sadness, anger, fear, disgust and surprise), not on the intermediate emotions (emotions that can be represented as the mixture of two primary 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.

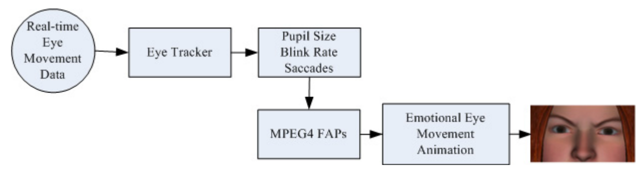


Figure 1: Framework of emotional eye movement synthesis framework

2. SYSTEM OVERVIEW

Our final research goal is to endow virtual agents with the emotional expression capability through eye movement, and employ it as one of the fundamental methods to convey emotion, on a par with facial expression, emotional speech synthesis and emotional gesture in virtual agents. The overall framework of our research is illustrated in Fig. 1. For pupil size, blink rate and saccade, we use eye tracker to capture and analyze real-time eye movement data, then compute statistics about the involved FAP [3]. The FAP file contains the values of cheek, nose, eyebrow, eyelid and eyeball that associate with the eye movement. Finally, a rule based method is employed to generate emotional eye movement for primary and intermediate emotions in virtual agents.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
ICMI-MLMI'09, November 2–4, 2009, Cambridge, MA, USA.
Copyright 2009 ACM 978-1-60558-772-1/09/11 ...$10.00.

3. REAL-TIME EYE TRACKER DATA

Research has demonstrated that the pupil size[4], blink rate [5] and saccade [6] 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 are not concentrated on emotional eye movement and employed in emotional eye movement generation in virtual agents. In our research, we employ the eye-tracker to capture real-time eye movement from subjects to extract the spatiotemporal characteristics of the eye movement for positive and negative emotions. Then, pupil, blink and saccade movement can be synthesized and coded into appropriate FAP values [3].

3.1 Participants and stimuli

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. Seventy-five pictorial stimuli, 25 per category were selected from the IAPS (International Affective Picture System)[7]. 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)[8]. 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 stimuli, and the upper end represented a very positive emotional stimuli. The center of the scale represented a neutral stimuli. All the stimuli were about 10s long.

3.2 Procedure

First, 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 an real-time eye tracking system SMI IVIEW.

3.3 Data analysis

3.3.1 Pupil size

In the experiment, the eye tracker recorded all the subjects' pupil sizes during the positive, negative and neutral stimuli. For the data during the stimuli, a 2×3 (gender $\times$ 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 $t = 0.3, df = 49, p > 0.5$. 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.

3.3.2 Blink rate

Blink has also been related to emotional reactions. We got the mean blink rate value of the 50 subjects for positive, negative and neutral stimuli. For the data during the stimuli, a 2×3 (gender $\times$ 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 $t = 0.6, df = 49, p > 0.5$. It also can conclude from the data that blink rate is linearly related to the valence of emotion.

3.3.3 Saccade

Saccade is rapid movement of eyes from one gaze position to another. Magnitude and direction are the conventions 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[9]. 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 1.

For the data during the stimuli, a 2 × 3 (gender × 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 $t = 7.3, df = 49, p < 0.001$.

4. SYNTHESIS RULE

After getting real-time 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 [10]. We utilize both the emotion wheel and the angular [11] measure to generate primary and intermediate emotional eye movement. According to the emotion wheel, emotions are distributed as a circle, we locate each emotion's position in the circle. 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 the X-axis by:

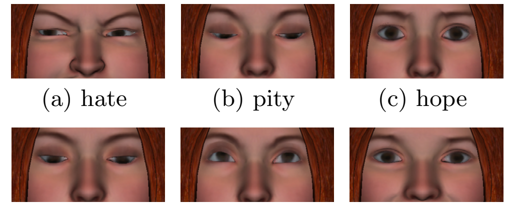
$$\alpha = arctan\frac{a - \overline{a}}{e - \overline{e}} \tag{1}$$

We will add pupil size, blink rate and saccade information to the eye movement. The rule is described as follows:

Pupil size:

$$Emotion_{pupil} = \frac{a}{\overline{a}}\overline{pupil} \tag{2}$$

where $\overline{pupil}$ depicts the mean pupil size we got in the section 3.3.1. The value is 6.1mm and 6.2mm for the positive and negative emotion respectively. $\overline{a}$ is the mean values of all values of activation 4.5. The $Emotion_{pupil}$ 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 (7.6mm for positive and 7.2mm for negative) or min value (4.6mm for positive and 5.0mm for negative) we got in the experiment, it will be set to the extremum value.



(a) hate (b) pity (c) hope (d) shame (e) pride (f) gloating

Figure 2: Some key frames of intermediate emotion eye movement based on our framework

Blink rate:

$$Emotion_{blink} = \frac{a}{\overline{a}}\overline{blink} \tag{3}$$

where $\overline{blink}$ denotes the mean blink rate. The value is 3.2 per-10s and 2.8 per-10s for the positive and negative emotion respectively as we got in the section 3.3.2. $\overline{a}$ and a are the same as in equation(2). The $Emotion_{blink}$ is the primary and intermediate emotions' blink rate we want to get. Note that if the value exceeds the max value (4.9 per-10s for positive and 4.5 per-10s for negative) or min value (1.4 per-10s for positive and 1.5 per-10s for negative) we got in the experiment, it will set to be the extremum value.

Saccade:

The rule for saccade generation is a modified version of the method described by Lee[9]. The saccade magnitude can be obtained from the function of equation(4)

$$A = -6.9 * log(P/15.7) \tag{4}$$

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. This guarantees 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 emotion state.

The direction is determined based on the distribution shown in Table 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, 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. (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.

Fig. 2 shows some key frames of the animated face with gaze behaviors for some intermediate emotions based on our rules employed MPEG-4 FAP.

5. EVALUATION

In this experiment, we evaluated the realistic generation of emotional eye movement. This was achieved through

Table 1: Distribution of saccade directions

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

Table 2: Percentage of recognition of primary and intermediate emotions

emotion	Per.1	Per.2	emotion	Per.1	Per.2	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						

the subjective evaluation of primary and intermediate emotions expression in a synthetic female face "Alice". First, fifty subjects were asked to judge twenty-two videos of emotion 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. Then, the twenty-two videos of facial expression generated by Raouzaiou's work were modified, replaced the eye movement related FAP values generated by our framework. 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 showed in Table 2, 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 had higher recognition rate than intermediate emotions. Some intermediate emotions had 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 were always recognized as joy. In general, the results were more or less expected and suggested that our framework worked well enough to bring measurable benefit to the human-agent interaction.

6. CONCLUSIONS

In this paper, we present a computational framework that enables virtual agents to express emotion through eye movement based on the parameters picked from the real-time eye movement data. We believe we propose 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 framework worked well enough to enhance the emotion expression in virtual agents.

7. ACKNOWLEDGEMENTS

This work is supported by the National Research Foundation for the Doctoral Program of Higher Education of China (No.20070006057) and National Nature Science Foundation of China (No.60873269).

8. REFERENCES

[1] E. Gu and N.I. Badler: Visual Attention and Eye Gaze During Multiparty Conversations with Distractions, Proceeding of International Conference on Intelligent Virtual Agents, 2007, pp.193-204

[2] B. Lance and S. Marsella: The Relation between Gaze Behavior and the Attribution of Emotion: An Empirical Study, Proceeding of International Conference on Intelligent Virtual Agents, 2008, pp.1-14

[3] MPEG4: Moving Pictures Expert Group, ISO/IEC 14496, International Standard on Coding of Audio-Visual Objects

[4] M.M. Bradley, L. Miccoli, M.A. Escrig and P.J. Lang: The pupil as a measure of emotional arousal and autonomic activation, Psychophysiology, 2008, vol. 45, pp. 602-607

[5] K. Takashima, Y. Miyata, Y. Itoh, Y. Kitamura and F. Kishino: Effects of avatar's blinking animation on person impressions, In Proceedings of Graphics interface, 2008, pp. 169-176

[6] J. Kissler and A. Keil: Look-don't look! How emotional pictures affect pro- and anti-saccades, Exp Brain Res, 2008, vol. 188, pp. 215-222

[7] P.J. Lang and M.M. Bradley and B.N. Cuthbert: Affective ratings of pictures and instruction manual, Technical Report A-8. University of Florida, Gainesville, FL.

[8] M.M. Bradley and P.J. Lang: International affective digitized sounds (IADS), Stimuli, instruction manual and affective ratings (Tech. Rep. No. B-2). Gainesville, FL: The Center for Research in Psychophysiology, University of Florida

[9] S. Lee, J. Badler and N. Badler: Eyes alive, ACM Transactions on Graphics, 2002, vol. 21, no. 3

[10] C. Whissel: The dictionary of affect in language, in Emotion: Theory, Research and Experience, Academic Press, 1989, pp. 113-131

[11] R. Plutchik: Emotion: A Psychoevolutionary Synthesis, Harper and Row, New York, 1980