Automatically Generating Eye Motion in Virtual Agents *

Paula S.L. Rodrigues
Departamento de Informática - PUC-Rio
paula@inf.puc-rio.br

Rossana B. Queiroz
PIPCA - UNISINOS
fellowsheep@gmail.com

Leandro Barros
PIPCA - UNISINOS
lmb@exatas.unisinos.br

ABSTRACT
Eye Motion has a fundamental role in verbal and non-verbal communication among people. It can express emotions and regulate the flow of conversation. For this reason, it is very important that these movements can be realistic in Computer Animation's applications, such as virtual humans in general and game/movies characters. This paper presents a model for producing eye movements in synthetic agents. The main goal is to improve the realism of facial animation of such agents during conversational interactions. The contribution of this work is the inclusion of saccade behaviours in eyes motion, correlated with emotional attributes and also dependent of agents' role during interaction, e.g. if agent is listening or talking. We investigate the impact of face animations in human-computer interactions by including saccade motions, emotional attributes and role's dependent animation. The experimental results indicate that increasing details in eyes animation can significantly improve the realism of facial animation.

Keywords
Eyes animation, facial animation, saccades.

1. INTRODUCTION
Virtual humans have been widely used in many applications. One is concerned with the interaction between humans and computers using ECA (Embodied Conversational Agents) [2], which are related with specific type of synthetic autonomous agent that aims to establish friendly dialogues with the user, enough to be utilized as system interfaces, entertainment applications, teaching tools, among others. In this case, the agents' reactions should help in user immersion as well as to provide a believable interaction. Other kind of application is related with games, movies and simulations, in which it is possible to observe synthetic agents, which interact with others agents, and not necessarily with users. Even in such type of application, the behaviours' realism is very important, to increase the credibility of character. In all these cases, once we are interested in improving the realism of agents interaction, the facial animation is a very relevant topic. The research in facial animation area has been very intense during many years. Several works have been proposed in order to animate facial muscles with speech or emotion. However, less attention has been paid to eyes animation, which poses a challenge for automatic animation systems, that aim to generate eyes animation based on dialogue or other kinds of input. Indeed, some eyes movements are activated in an unconscious way, for instance while we talk, our mind is flooded by the meaning of the words, which are metaphorically linked with another signs and so on, and we search for information in such senses beyond our intellect to formulate what we want to say. After we have formulated what we wanted to say, the unconscious takes the control of our expressions. Not even an experienced orator has idea of all kinds of gestures or facial expressions one can perform during a speech [2]. Even actors, specialists of the artistic representation of human emotions, suffer of this unconscious behaviour. One example of such unconscious eye movement is the saccade motion, which is denoted by a continuous, stepwise manner of movement as opposed to a fluent and continuous one [1].

The main goal of this paper is to provide a model for automatic eyes animation, including saccade behaviours associated with emotional attributes and also dependent of agents' role during interaction, e.g. if agent is listening or talking. Our model is in some way inspired by two works. First of all, the system ResponsiveFace [7] which provides a parametrized facial animation based on muscular deformation. Second, our model is also based on work proposed by Lee et al. [6] which describes a statistical model for saccade behaviour in eyes motion. The results presented in this paper aim to evaluate the impact of having saccade behaviour, associated with emotional attributes and roles-dependent animation. To achieve this goal, we used the criterion of visually appreciation of facial animations generated with our prototype.

The remainder of this paper is organized as follows. Some methods for eyes animation are discussed in Section 2. Section 3 describes our system for automatic eyes animation, while Section 4 describes our facial animation system that includes emotional state, head motion and character's role

*(Produces the permission block, copyright information and page numbering). For use with ACM_PROC_ARTICLE-SP.CLS V2.6SP. Supported by ACM.

during a performance, Section 5 discusses some experimental results obtained with our prototype by changing parameters in facial and eyes animation. Finally, our conclusions and final remarks are drawn in Section 6.

2. RELATED WORK

Research on facial animation has not deeply focused on eyes motion, even if everybody agrees that eyes play an essential role in inter-personal communication. Moreover, several systems have proposed face-to-face conversation with a user. Poggi et al. [8] focused their work on gaze behaviour and propose a meaning-to-face approach, allowing to simulate automatic generation of face expressions based on semantic data. Like any communicative signal, gaze includes two different aspects: a signal, which encompasses a set of physical features and the dynamic behaviour of eyes in gaze; and the meaning which is related with a set of beliefs that gazes communicates.

Fukayama and collaborators [4] propose a gaze movement model that enables an embodied interface agent to convey different impressions to users. Managing one's own impression, in order to influence the behaviours of others, plays an important role in human communications. With this purpose, they build the gaze movement model based on three parameters selected from psychological studies: amount of gaze, mean duration of gaze and gaze points while averted. In their work, they also present experiments in which subjects evaluated their impressions created by gaze patterns produced by altering the gaze parameters. The results indicated that reproducible relations exist between the gaze parameters and impressions.

Cassel et al. [3] address the problem of designing conversational agents that exhibit appropriate gaze behaviour during dialogues with human users, mainly focused on the relationship between information structure and gaze behaviour. A behavioural model used to simulate realistic eye-gaze behaviour and body animations for avatars in a shared immersive virtual environment was proposed by Vinayagamoorthy et al. [9]. Their model was based on data and studies carried out on the behaviour of eye-gaze during face-to-face communication. Also, Garau et al. [5] investigate the importance of gaze behaviour in avatars representing people engaged in conversation. They compared responses in four mediated conditions: video, audio-only, and two avatar conditions.

In our work, differently from the previous cited papers, the main goal is to evaluate the visual impact of saccadic eye movements in virtual humans, when they are connected with emotional and roles-dependent animations. The saccadic eye movements are related with the rapid motion which exists in both eyes between two gaze positions. Lee et al. [6] present a statistical eye movement model, which is based on both empirical studies of saccades and acquired eye movement data. This work is used as basis for our facial animation model in order to provide the saccade behaviour.

Our model, which includes the saccade motion, emotional attributes and different roles during the conversational process, is developed as part of ResponsiveFace [7]. Next sections present details of our animation architecture.

3. EYES ANIMATION MODEL

Eye motion in our system is based on work proposed by Lee et al. [6] which describes a statistical model for saccade behaviour. Indeed, they use eye trackers in order to extract the dynamic characteristics of the eye movements. More specifically, they were interested in extracting saccadic behaviour by obtaining the image (x, y) coordinates of the pupil center. The main components of Lee model are Attention Monitor, Parameter Generator and Saccade Synthesizer. The first one is responsible for monitoring the system, also detecting aspects such as: subject role in the conversation (talking or listening mode), head rotation, as well as to detect whether the current frame has reached the gaze duration. If the direction of head rotation changed, and its amplitude is bigger than a specified threshold, or if the timer for gaze motion is expired then this component activates the component Parameter Generator. Such component, responsible for generation of eyes parameters, is used to determine the saccade magnitude, direction, duration and instantaneous velocity. Then, the Saccade Synthesizer is activated.

In our work, we based the generation of our saccade behaviours in equations proposed by Lee et al. [6], and also Vinayagamoorthy et al. [9].

Firstly, as proposed by Lee, we generate a random number between 0 and 15, which corresponds to percentage of occurrence of saccadic motion P. Then, magnitude of saccadic motion can be obtained by Equation 1,

$$A = -6.9 \log(P/15.7) \quad (1)$$

where A is saccade magnitude in degrees. Also, Lee defines that the maximum saccade magnitude is limited to 27.5 degrees for talking mode, while 22.7 degrees for listening mode. The saccade direction can follow the head rotation, whether a threshold is attained. On the other hand, the direction is determined based on eight specific directions, as detailed in [6]. Concerning saccade duration, Equation 2 is used with values $d = 2.4$ ms/degree and $D_0 = 25$ ms.

$$D = D_0 + dA \quad (2)$$

Equations 1 and 2 are used as proposed by Lee et al. [6]. However, the saccade velocity motion is used as proposed by Vinayagamoorthy et al. [9] (Equation 3), where authors improve the referred equation, avoiding negative values.

$$V = 14e^{[-(\pi/4)(X-3)^2]} \quad (3)$$

where V is instantaneous velocity (degrees/frames). X is related with time in frames of a saccadic motion, which happens into an interval of Y as described next. Lee described that in talking mode, the inter-saccadic interval is generated by defining 2 classes of gaze, mutual and away. The equations 4 and 5, respectively define such situations,

$$Y = 2.5524 \times 10^{-4} X^2 - 0.1763X + 32.2815 \quad (4)$$

$$\begin{aligned} Y = &1.8798 \times 10^{-5} X^4 - 0.0034X^3 \\ &+0.2262X^2 - 6.7021X + 78.831. \end{aligned} \quad (5)$$

In listening mode, the inter-saccadic interval is obtained by choosing a random gaussian number using average and standard deviation proposed in [6].

The main contribution of this paper in the context of saccade behaviour is the inclusion of such model in the facial animation system, also integrating it in the roles dependent animation and emotional status. In the next section, we give an overview of our facial animation model and show how we integrate the saccadic behaviour into the system.

4. FACIAL ANIMATION SYSTEM

Our facial animation model is based on the ResponsiveFace developed by Ken Perlin [7]. The goal of this system is to create an embodied agent that reacts with responsive facial actions in real-time with convincing emotive expressiveness.

One of the main characteristics of the ResponsiveFace is that it promotes a "minimalist approach" for experiments in computer animation. First, the virtual agent has been designed with a non-photorealistic look. Second, it defines a minimal number of facial expression elements that can produce a convincing impression of character and personality. In this way, the system emphasizes the essential aspects for depicting emotion in a virtual character.

The implementation of the system is summarized as follows: the face geometry is represented by a 3D polygonal mesh and rendered using a cartoon shading style (see Figure 1).

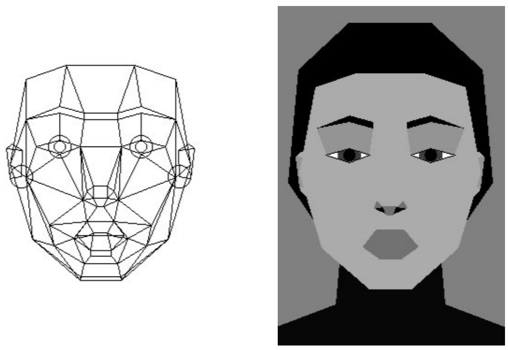

Figure 1: Polygonal Face Model and Cartoon Style Shading

The face animation is controlled by pseudo-muscles that act on sets of vertices of the face geometry. Each pseudo-muscle gives control with a low-level degree of freedom for the facial expression. These controls are called *Flexors* and are parametrized in the interval $[-1, 1]$, that is from full contraction (-1) to relaxed (0), to full extension (1).

Expressions are created by combining the action of groups of flexors into higher-level controls or *Actions*, that vary in time. The system is designed to allow mixing of these high-level actions using a composition of movements that blend layers of actions in natural ways for a final expression animation.

A basic set of emotions can be modeled using these expressions. Snapshots are shown in Figure 2, respectively they express emotional states such as: surprised, happy, arrogant, angry, disappointed, annoyed, and frightened.

We have extended the ResponsiveFace system in several ways: 1) synchronized audio reproduction; 2) lip-sync for speech animation; 3) use of multiple procedural motion generators; and 4) input parameters to control animation.

The system has as input an audio file in the wave format, and parameter files that are interpreted by motion generators. The output is an animation of the virtual character with layered expressions and synchronized with the audio.

The animation pipeline can be divided into two phases: a pre-processing stage and a real-time animation reproduction. In the pre-processing stage, the audio and parameter files are read and validated. Each parameter file defines a different group of transitions (or actions) during the time. We are working with four types of parameter files: emotion transitions, eye gaze movements, head movements and talking/listening status transitions. The parameter file has also a general format and can be customized for different types of parameters (i.e., binary, scalar, vector, etc.). Furthermore, the format encodes a time varying parameter. The general structure of these files is as follows: a header describing the parameter elements; and a list of time intervals with associated parameter values. As an example, a head motion file is shown below:

```
header
# head motion
element transition 2
property int duration
property double turn_value
property double nod_value
property double tilt_value
end_header
2;1;0.8;0
3;-1;-1;0.7
```

Once a parameter file has been loaded, each element transition is transformed by the corresponding motion generator in a common data structure for the animation, which we call *script animation element*. This structure stores important values that will be used during the real time animation reproduction stage like initial time, duration, the specific value for each flexor (or muscle) and the listening/talking status (the character role during the conversational process). Flexors that do not belong to the element transition receive a "do not care value". In the example, when converting the head movement to a script animation element the eyes flexors receive the "do not care value".

Motion layering is obtained by the subsequent activation

Figure 2: Snapshots of possible emotional states defined in ResponsiveFace System.

of different motion generators. Currently we are using the motion generators, mostly to control head motion, face expressions and eye motion. The saccade behaviour described in the previous section is implemented as a motion generator for the experiments described in Section 5.

The output of the preprocessing stage is a temporal ordered list with all script animation elements. In the animation reproduction stage, the system plays the audio and displays the animated face, synchronized in real-time. One thread is responsible for playing the audio and another thread synthesizes the motion and renders the virtual character's face based on the script animation specification.

We have also implemented an interactive front-end for the system in order to allow the user to perform different experiments and format applications, shown in Figure 3.

The user can interact directly with the virtual character or read in a script with the parameters of an animated performance.

In the *interactive mode* (see the right side of the Figure 3), the user can select different emotional states, change the associated expressions and apply different visualization controls for the animation.

In the *batch animation mode* (see the left side of the Figure 3), the user specifies parameter files associated with an animation. The architecture of the system is general enough that various motion generators and parameters can be added into the system.

So far, we have included the following parameters that are used for our experiments: 1) recorded audio; 2) emotion state; 3) eye gaze motion; 4) head movement; 5) speech action.

The recorded audio is used for playback. The emotion state is used by the expression generator. The eye gaze motion and head movement are used by the saccade generator. The speech action state is used by both expression and saccade generators.

Next section presents some results of using the system to investigate the influence of eye movement in the expressiveness of a virtual character.

5. RESULTS

This section presents experimental results we have obtained by using our facial animation system. We applied a subjective test with a group of people to evaluate the importance of our system eyes movement support for facial animation with speech synchronization. In such experiments, we played the animation using cartoon style rendering and including the features of head movement, different facial expressions based on emotions, and character role during the conversation (i.e. listening or talking).

In order to focus on the eyes animation, we decided to display only the upper region of the face (i.e., the eyes and eyebrows). The animation was played for two specific dialogs that establish the emotional context. The first dialog was extracted from Shrek 2 movie where our character acts like the woman (Fiona) talking with Shrek. The second dialog is a part of Ice Age 2 film and our actor plays the female mammoth role. Both movie tracks were chosen because they present conversational scenes of known animation films and with a great emotion transition during the dialog between the characters.

Table 1 shows the dialogue for the first movie to exemplify the input information for generating the final facial animation.

To realize our experiments, we generated three video versions of our character animation for each movie dialog:

a) with no eyes movement;

b) with eyes movement based only on Perlin noise; and

c) with eyes movement based on saccadic behaviour (blink motion is based on Perlin noise).

In case (a), the character eyeballs remain fixed. The character has movements only due to the emotion transitions and head movements during both dialogs.

In case (b), we applied the noise function developed by Ken Perlin [7]. Indeed, this noise function allowed computer graphics artists to better represent the complexity of natural phenomena in visual effects for the motion picture industry. More focused on generating natural-looking textures, Perlin noise has also been applied in other applications, such as in the development of ResponsiveFace [7]. In case (b), we

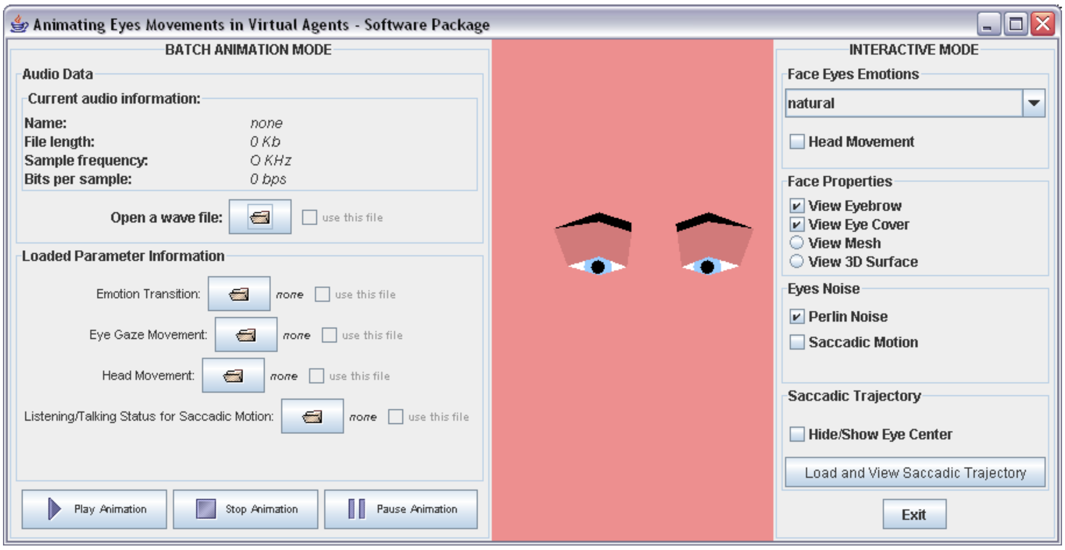

Figure 3: Interactive front-end for experimental set-up (the left side is the 'Batch Animation Mode' and the right side is the 'Interactive Mode').

Table 1: Simulated Dialogue: Information noted in XML file.

Simulated Dialogue			
Speaker	Character Speech	Fiona's role	Fiona's Emotion
Fiona	Very nice, Shrek	Speaking	Half Angry
Shrek	What? I told you coming here was a bad idea.	Listening	Natural
Fiona	You could've at least tried to get along with my father.	Speaking	Angry
Shrek	I don't think I was going to get Daddy's blessing, even if I did want it.	Listening	Natural
Fiona	Do you think it might be nice if somebody asked me what I wanted?	Speaking	Half Angry
Shrek	Sure. Do you want me to pack for you?	Listening	Natural
Fiona	You're unbelievable!	Speaking	Surprised
Fiona	You're behaving like a...	Speaking	Half angry
Fiona	(Silence)	Silence	Disappointed (2 seconds)
Shrek	Go on! Say it!	Listening	Natural
Fiona	Like an ogre!	Speaking	Angry
Shrek	Here's a news flash for you! Whether your parents like it or not... I am an ogre!	Listening	Natural
Shrek	ARGHHHHHHHHHHHHHHHHHHHH!	Listening	Frightened
Shrek	And guess what, Princess? That's not about to change.	Listening	Half Angry
Fiona	(Silence)	Silence	Natural (11 seconds)
Fiona	I've made changes for you, Shrek. Think about that.	Speaking	Disappointed

only applied Perlin noise for generating the eyes movement. Figure 4 on left side, shows the trajectory of pupil center as a function of time. Time increases along the figure height (the beginning of simulation is located at the bottom of the image, and the last frame is in its top).

In case (c) we applied saccadic behaviour for the animation production. Saccadic behaviour was used to generate the pupil movement, however, the blink motion was still generated as a function of Perlin noise. Future investigations will focus on the blink motion more adapted to the saccadic model. For saccadic behaviour we used basically the model proposed by Lee et al. [6], as described in Section 3. The right-side of Figure 4 shows the pupil trajectory generated based on saccadic behaviour during 5000 frames, similar to the case (b).

We can clearly note in such graphics that Perlin noise behaves as expected, i.e. providing a random motion of pupil center. On the other hand, saccadic motion presents the real motion of pupil center based on statistical data. It describes natural and involuntary eye motion, escaping from the eye center and returning to such position in certain periods of time.

For case (a), we did not consider different roles during the conversation, what also happened in the case (b), since Perlin noise does not deal with such aspect. On the other hand, saccadic behaviour (case c) generates different motions based on the conversational role. Figure 5 shows some snapshots of the generated animation for the first dialog (Shrek 2) using the case (c).

The subjective test we conducted is similar to the one applied by Lee et al. [6], but we added some questions (Q5 and Q6) specifically related with our work. The six questions are

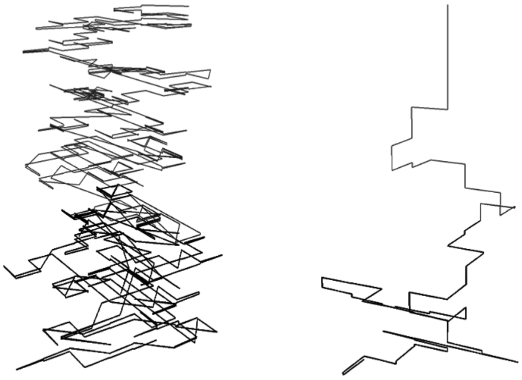

Figure 4: On left-side: Pupil motion in function of time, using Perlin noise. On right-side: Pupil motion also in function of time, using saccadic motion.

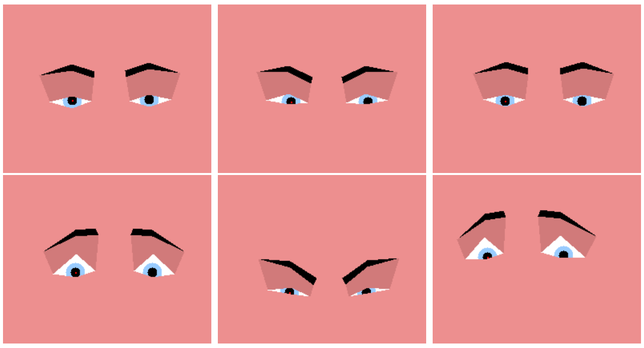

Figure 5: Snapshots of our simulation using cartoon style rendering. The snapshots refer respectively to the second and first speech of Fiona, second speech of Shrek, fourth and seventh speech of Fiona and sixth speech of Shrek, as detailed in Table 1.

described as follows:

- Q1: Did the character on the screen appear interested in (5) or indifferent (1) to you?
- Q2: Did the character appear engaged (5) or distracted (1) during the conversation?
- Q3: Did the personality of the character look friendly (5) or not (1)?
- Q4: Did the face of the character look lively (5) or deadpan (1)?
- Q5: Did the eyes character transmitted emotion (5) or were without emotion (5)?
- Q6: In general, which score do you give to the character performance in those dialogs? (5) Good or (1) bad?

Fifty one people were invited to participate in our test. Each one gave a score, varying from 1 to 5, for our six questions. Note that the higher scores correspond to more positive attributes in a conversational partner.

To answer these questions, the participants saw six different video sequences: the three different cases aforementioned for the two selected animation movies.

Figure 6 and Figure 7 summarize the average score for each one of our six questions applied for Shrek 2 and Ice Age 2, respectively. These graphics indicate that the three character eyes movement types have significant different mean scores.

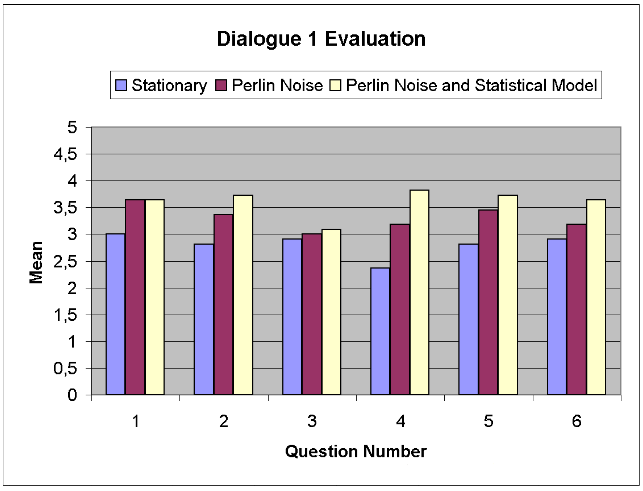

Figure 6: Results of subjective evaluations for dialogue 1 (Shrek 2 animation). The graphic shows the average score for each question.

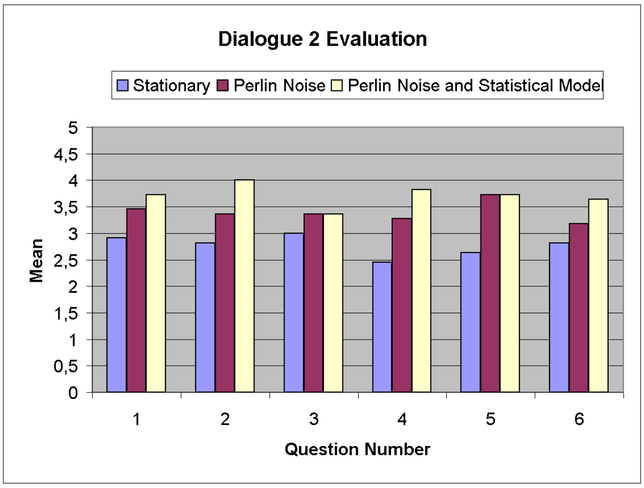

Figure 7: Results of subjective evaluations for dialogue 2 (Ice Age 2). The graphic shows the average score for each question.

Analyzing the graphics, we observe that for all questions and for both movies the mean score for saccadic movement is never worse than the mean score for the other cases. Specifically in the answers for question 1, we perceive that the absence of eyes movement contributes for a worse character performance evaluation, as expected.

For questions 2, 4 and 6 we identify that applying saccadic motion the improvement in the character performance, comparing with the exclusive use of Perlin Noise, has been greater (14.86%, 18.31%, 14.29%, respectively, of improvement).We understand that since questions 3 and 5 are more related with emotional aspects of the character, in these cases the pupil movement is more a complement for the facial expression. However, when the evaluation concerns specifically with the emotion transmitted through the eyes (question 3), the pupil movement (whether saccadic or random) enriches the user experience.

We conclude with the experimental results that it is necessary to combine a noise with the saccadic motion, but not a pure random noise, since it is important to consider the character emotion and other context information like if the character is listening or talking during the conversation. The noise incorporation aims at providing a more natural behaviour, influencing in the pupil motion rhythm. Actually, several participants highlighted the performance of case (c) because of the character kept their attentional focus in the conversation but at the same time presented a natural behaviour when deviating its look.

Finally, we use a realistic face manually animated based on the motion generated with our model. The goal is to improve the visual realism, since we believe that the impact of including saccadic behaviour can be even better if applied to a realistic face. Figure 8 shows some snapshots of such animation.

The main goal of this section was to show (also by analyzing the short movies which accompany this paper) that the inclusion of emotion and saccadic behaviours increase the realism of facial animation. Moreover, if the character is to be used to provide inter-communication with users, each visual detail that could be improved in facial animation, certainly contributes to the feeling of immersion, what is very desirable in such type of system. Next section presents final remarks and also future directions we intend to go forward.

6. FINAL REMARKS

In this paper we investigated the influence of eye motion in the expressiveness of a virtual character. To this end, we improved the statistical model for saccade behaviour of Lee et al. [6] and combined it with dynamic expressions of the upper face in the eye region. We implemented these mechanisms as motion generators in a testbed system based on the ResponsiveFace [7]. We also produced several experiments to determine the impact of natural eye motion and expressive reaction in a conversation with emotional content. The preliminary results from these experiments indicate that only through the combination of different mechanisms it is possible to achieve a high level of expressiveness in a virtual character. Moreover, the eyes play a leading role for directing the characters face in a conversation and also revealing emotional state.

As for future work, we would like to develop an statistical model for eye blink motion, that is integrated with the saccadic movement. Another avenue for future research is the use of situated eye gaze behaviour combined with emotional and contextual information from a multi-agent conversation.

7. ADDITIONAL AUTHORS

Additional authors: Bruno Feijó (Departamento de Informática PUC-Rio, email: `bruno@inf.puc-rio.br`), Luiz Velho (IMPA, email: `lvelho@impa.br`) and Soraia R. Musse (Faculdade de Informática PUCRS, email: `soraia.musse@pucrs.br`).

8. REFERENCES

[1] W. Becker. *In The Neurobiology of Saccadic Eye Movements*. R H Wurtz and M E Goldberg (eds), 1989.

[2] J. Cassel, J. Sullivan, S. Prevost, and E. Churchill. *Embodied Conversational Agents*. MIT Press, New York, 2000.

[3] J. Cassell, O. Torres, and S. Prevost. Modeling Gaze Behavior as a Function of Discourse Structure. In *First International Workshop on Human-Computer Conversations*, Bellagio, Italy, 1997.

[4] A. Fukayama, T. Ohno, N. Mukawa, M. Sawaki, and N. Hagita. Messages embedded in gaze of interface agents — impression management with agent's gaze. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 41–48, New York, NY, USA, 2002. ACM Press.

[5] M. Garau, M. Slater, S. Bee, and M. A. Sasse. The impact of eye gaze on communication using humanoid avatars. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 309–316, New York, NY, USA, 2001. ACM Press.

[6] S. P. Lee, J. B. Badler, and N. I. Badler. Eyes alive. In *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 637–644, New York, NY, USA, 2002. ACM Press.

[7] K. Perlin. Layered compositing of facial expression. In *ACM SIGGRAPH - Technical Sketch*, 1997.

[8] I. Poggi and C. Pelachaud. Signals and meanings of gaze in animated faces. In *Language, Vision and Music*, pages 133–144. Mc Kevitt, Paul, Sen Nuallin and Conn Mulvihill, 2002.

[9] V. Vinayagamoorthy, M. Garau, A. Steed, and M. Slater. An Eye Gaze Model for Dyadic Interaction in an Immersive Virtual Environment: Practice and Experience. In *Computer Graphics Forum*, volume 23, 2004.

Figure 8: Snapshots using realistic face animated for similar frames than the illustrated in Figure 5.