Environment Expression: Telling Stories through Cameras, Lights and Music

Celso de Melo, Ana Paiva

IST- Technical University of Lisbon and INESC-ID
Avenida Prof. Cavaco Silva – Taguspark
2780-990 Porto Salvo, Portugal
{celso.de.melo,ana.paiva}@inesc-id.pt

Abstract. This work proposes an integrated model – the *environment expression model* – which supports storytelling through three channels: cinematography, illumination and music. Stories are modeled as a set of *points of interest* which can be characters, dialogues or sceneries. At each instant, audience's focus is drawn to the highest priority point of interest. Expression channels reflect the type and emotional state of this point of interest. A study, using a cartoon-like application, was also conducted to evaluate the model. Results were inconclusive regarding influence on story interpretation but, succeeded in showing preference for stories told with environment expression.

1 Introduction

Digital technology progress inspires replication of the kind of storytelling seen in the arts. In theatre, a story is told using drama, sceneries, makeup, lights and sound. In cinema, the camera introduces a new expression channel. In digital worlds, besides similar channels, synthetic characters and virtual environments bring new potential.

This work is about the role of virtual environments in virtual storytelling. In concrete, an integrated model is proposed which supports storytelling through three channels: cinematography, illumination and music. Each channel reflects differently, at each instant, the character, dialogue or scenery which is being given focus.

Regarding the rest of the paper, section 2 describes the model and channels, section 3 presents an evaluation of the model and, finally, section 4 draws some conclusions.

2 Environment Expression

Stories, which can be scripted or interactive, are modelled as a set of *points of interest* which compete for the audience's attention. These can be of three kinds: (1) *characters*; (2) *dialogues*; (3) *sceneries*, i.e., the story's setting itself. Each point of interest is associated with a priority which can change through the course of the story.

The *environment expression model* receives as input a story. The story is told through the director and three expression channels. The *director* focuses, at each

instant, the audience's attention to the highest priority point of interest. Then, the three *environment expression channels – cinematography*, *illumination* and *music* – present it to the audience's senses. Fig.1 summarizes this model.

Emotion synthesis and expression are also explored in this work. Regarding synthesis, a character has an *emotional state* based on the OCC emotion theory [1]. Dialogues' and sceneries' emotional states reflect the participants' emotional states. Regarding expression, emotion is expressed through the expression channels. More information on this work's emotion synthesis and expression is found in [2].

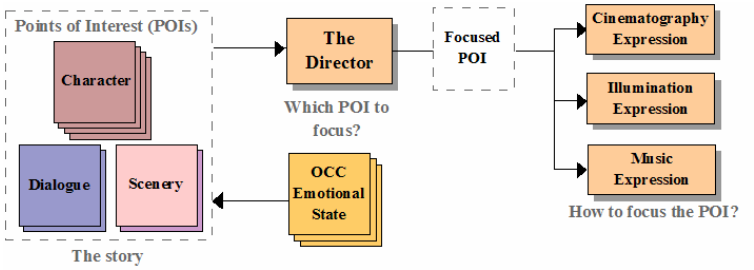

Fig. 1. The environment expression model

2.1 Cinematography

According to [3], camera shots vary according to *distance* – extreme close up, close up, medium, full shot and wide shots - and *movement* – point of view and establishment shots. Distance influences the audience's emotional attachment to the point of interest. Point of view shots convey the point of interest's perspective. Establishment shots overview the setting thus, contextualizing the action. Regarding shot sequences, it is important not to cross the line of action to avoid disorienting the audience. In this sense, for dialogues, the *triangle principle* is widely used.

In this work, the cinematography channel selects the shots according to the focused point of interest's type as follows: (1) if it is a character, a full shot or point of view shot is randomly chosen; (2) if it is a dialogue, the appropriate triangle principle shot is applied according to who is talking; (3) finally, if it is a scenery, establishment shots are chosen. To avoid confusing the audience, camera shots can change only after a certain amount of time has elapsed. Finally, shot distance and angle vary according to the focused point of interest's emotional state [2].

2.2 Illumination

According to [4], the *three-point lighting* technique, widely used to focus characters, is composed of the following lights: (1) *key light*, which is the main source of light; (2) *fill light*, which is a low-intensity light that fills an area that is otherwise too dark; (3) *back light*, which is used to separate the point of interest from its background.

In this work, the illumination channel uses a variation of the three-point technique to illuminate the focused point of interest. The key light is a point light placed between the point of interest and the camera along the 'point of interest-camera'

vector. Because it has attenuation and it need not focus a specific target, a point light was preferred to directional and spotlights. Key light attenuation and color varies according to the point of interest's emotional state [2]. The fill light role is assumed by general scene illumination, responsible for illuminating the rest of the scene, and back lights were ignored for their low visual impact and performance reasons.

2.3 Music

According to [5], the relation between music and emotion varies according to: *Structural features* – which relate the music's structure with emotions; *Performance features* – which refer to the influence of the *interpretation* of the music. Regarding the first, tempo is one of the most influencing factors affecting emotional expression in music. Fast tempo may be associated with happy/exciting emotions and slow tempo with sad/calmness emotions. Regarding performance features, the expressive intention of the performer is converted into various cues during the performance.

In this work, the music expression channel reflects the focused point of interest's *mood valence* – positive, neutral and negative. To convey mood valence, music, with the same valence, is randomly selected from a library. To fill the library music was selected according to the following simple criteria: (1) Positive songs should have fast tempo and, if applicable, should have positive lyrics; (2) Neutral songs should have medium tempo; (3) Sad songs should have slow tempo and, if applicable, should have negative lyrics. Regarding the association between lyrics' and music's valence, as the performer conveys the music's mood through cues, it is reasonable to expect that the lyrics' mood propagates to the performance's structural features.

3 Study

To test the model the *dancing solids* application was developed. This is a cartoon-like application with a simple story: "Once upon a time, there was a bunch of dancing solids. There were pyramids, cylinders and ellipsoids. There were boys and there were girls. The girls allured the boys. If the boy liked the girl, he would court her with a dance. If they both liked each other they simply got married. The end". The cartoon paradigm was chosen because it facilitates emotion expression which is essential to the story's plot. Furthermore, to force the audience to rely on environment expression for interpretation, characters were given simple bodies – geometric solids with eyes.

A study was conducted on this application to assess the relevance of environment expression on the audience's interpretation of the story plot, as well as the adequacy of this work's approach for each expression channel. The study, which was fully automated, was organized into four parts: (1) *Subject Profile* – where the subject's profile was assessed; (2) *Emotion Perception* - where the subject was asked to guess which emotions were being expressed by dancing solids using varying configurations of the expression channels; (3) *Music emotional valence* – where the subject was asked to classify 12 music compositions according to one of the following mood valences: positive/happy; neutral; negative/sad; (4) *Stories interpretation* – where the subject was told it would see two different stories when, in fact, it was presented with

two different versions of the *same* dancing solids story. Stories were assigned randomly a happy – girl marries boy – or unhappy ending – girl doesn't marry boy. Version A had no environment expression, while version B had all three channels active. After visualizing each version, the subject was asked whether the girl liked the boy. Additionally, before finalizing, it was asked which was the preferred version.

The study was presented to 50 students at Technical University of Lisbon. Subject average age was 23 years and 84% were males. Average time per inquiry was 12 minutes. Regarding emotion perception, results showed that environment expression, in particular the illumination channel, positively influenced emotion perception [2]. Regarding music emotion valence, average subject classification matched predictions for 92% of the music thus, revealing that mood valenced music selection based on tempo and lyrics emotional valence is sufficient to produce satisfactory results. Regarding stories interpretation, in general, subjects perceived correctly the stories' ending independently of environment expression (94.3% for version A and 82.9% for version B). This doesn't mean that environment expression has no influence on interpretation but, that the dancing solids, even though having simple bodies, were expressive enough to tell the whole story. Finally, regarding story preference, when the ending was unhappy both versions were equally enjoyed (60% of the subjects) followed by version B (35%). When the ending was happy, version B was preferred (50% of the subjects) followed by version A (33%). Thus, it seems that even though the study didn't succeed in showing the influence on story interpretation, it succeeded in showing environment expression relevance for the whole storytelling experience.

4 Conclusions

This works proposes a model for storytelling through three channels: cinematography, illumination and music. Stories are modeled as prioritized points of interest which can be characters, dialogues or sceneries. During storytelling, at each instant, the highest priority point of interest is focused differently by each of the expression channels.

Evaluation of this work revealed that: mood valenced music selection based on tempo and lyrics emotional valence is sufficient to produce satisfactory results; subjects perceived the story independently of environment expression; subjects preferred a version of a story told with environment expression to one which did not.

References

1. Ortony, A., Clore, G., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press (1988)
2. de Melo, C., Paiva, A.: Environment Expression: Expressing Emotions through Cameras, Lights and Music. Submitted to the 1st Affective Computing Conference (ACII05) to be held in Beijing, China
3. Arijon, D.: Grammar of Film Language. Silman-James Press (1976)
4. Birn, J.: [digital] Lighting & Rendering. New Riders (2000)
5. Juslin, P., Sloboda, J.: Music and Emotion: theory and research. Oxford University Press (2001)