# Environment Expression: Expressing Emotions 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.** Environment expression is about going beyond the usual Human emotion expression channels in virtual worlds. This work proposes an integrated storytelling model – the *environment expression model* – capable of expressing emotions through three channels: cinematography, illumination and music. Stories are organized into prioritized *points of interest* which can be characters or dialogues. Characters synthesize cognitive emotions based on the OCC emotion theory. Dialogues have collective emotional states which reflect the participants' emotional state. During storytelling, at each instant, the highest priority point of interest is focused through the expression channels. The cinematography channel and the illumination channel reflect the point of interest's strongest emotion type and intensity. The music channel reflects the valence of the point of interest's mood. Finally, a study was conducted to evaluate the model. Results confirm the influence of environment expression on emotion perception and reveal moderate success of this work's approach.

## 1 Introduction

The advent of digital technology has introduced several new ways to tell a story. Storytelling has evolved into a complex process involving sophisticated virtual characters capable of body, facial and voice expression and sophisticated virtual environments capable of cinematography, illumination and music expression.

This work is about virtual environments telling stories and expressing emotions. The idea behind environment expression comes from theatre. Theatre is one of the most complete forms of expression. Dramatic expression, text, sceneries, lights, make-up, sound, music and dance work together to tell a story [1]. With the advent of movies, new expression channels came to be, being the camera the most pervasive one. With the advent of digital technology, yet new channels of expression were created making it easier to break the rules of Nature.

This work proposes an integrated storytelling model – the environment expression model – capable of expressing emotions through three different channels: cinematography, illumination and music. The story is organized according to prioritized points of interest which can be either characters or dialogues. Characters synthesize cognitive emotions based on the OCC emotion theory. Dialogues have

collective emotional states which reflect their participants' emotional states. During storytelling, at each instant of time, the highest priority point of interest is focused differently by each of the environment expression channels.

The rest of this paper is organized as follows. Section 2 overviews the environment expression model. Section 3 describes the OCC based emotion state for both kinds of points of interest. Sections 4 to 6 describe, respectively, the cinematography, illumination and music expression channels. Section 7 describes a study conducted to assess the influence and relevance of environment expression in storytelling, as well as the adequacy of this work's approach. Finally, section 8 draws some conclusions.

## 2    Environment Expression Model

The *environment expression model* has the following components: (1) the *story module*; (2) the *director*; (3) and the three environment expression channels – *cinematography*, *illumination* and *music*. Fig.1 summarizes this model.
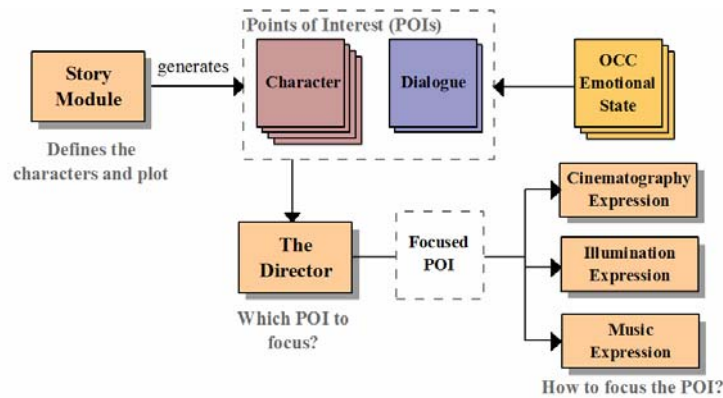


**Fig. 1.** The environment expression model

The story module essentially defines the story's plot and characters. At any given instant, a story can be defined by a set of *points of interest* which compete for the audience's attention. A point of interest can be a *character* or a *dialogue* between two characters. Furthermore, this module assigns priorities to the points of interest.

Usually in stories, characters perceive, synthesize and express emotions. This work considers cognitively generated emotions and, in this sense, uses the OCC emotion theory. The character's and the dialogue's emotional states are described in section 3.

One principle which is extensively used in theatre, animation and cinema is to focus the audience's attention to a single aspect of the story at a time [1]. This makes the message clearer. Using this principle, from all the generated story points of interest, the director focuses the audience's attention to the highest priority one.

Finally, an environment expression channel is a means by which the focused point of interest is presented to the audience. Besides just making the point of interest accessible to the audience's senses, these channels also express emotions. Each of the three explored channels is described in sections 4 to 6.

# 3     Emotion Synthesis

In this work characters can synthesize cognitive emotions. Emotion synthesis is based on the OCC emotion theory. Dialogues have collective emotional states which reflect their participants' emotional states. Following, subsection 3.1 overviews the OCC emotion theory, subsection 3.2 presents the character's emotional state model and subsection 3.3 presents the dialogue's emotional state model.

## 3.1     Background

The Ortony, Clore and Collins (OCC) emotion theory defines emotions as "valenced reactions to events, agents, or objects, with their particular nature being determined by the way in which the eliciting situation is construed". Thus, emotions result from cognitive interpretation of some emotion eliciting situation. The theory proposes 22 different *emotion types*, as well as a set of global and local variables [2]. As a general theory for emotions, however, it is incomplete. Though it proposes a mechanism for converting eliciting situations into cognitive emotions, not much is said on converting eliciting situations into the proposed variables' values or on emotion expression.

## 3.2     Character's Emotional State

The character's emotional state is based on a full implementation of the OCC theory, including its 22 emotion types, local and global variables. This work also explores emotion issues such as *decay, potential calculation, intensity reinforcement* and the effect of global variables on potentials, all of which are not solved by the OCC theory. As suggested in [2], only emotions whose potential is greater than a threshold are active. As suggested in [3], intensities are constrained to the interval [0; 10].

In Nature, an active emotion does not stay active forever as it decays with time [4]. Thus, for all emotion types,  decay is represented by function (1), based on [4], where $\Delta t$ is the time elapsed since the emotion was last elicited, $d$ is the emotion *decay rate* which is empirical, and $i_0$ is the intensity at the instant it was last elicited:

$$decay(\Delta t, d) = i_0 \times \exp(-0{,}1 \times d \times \Delta t) \ . \tag{1}$$

Potential is a function of local and global variables. The latter shall be addressed below. As suggested in [3], all emotion potentials are constrained to the interval [0; 10] and all local variables values to the interval [-10; 10]. Essentially, potential is a function of the eliciting situation which is defined as the values assigned to the local variables. As this assignment is not defined in the OCC theory, intuition was used. According to the OCC theory, different sets of variables are considered for different emotion types. Thus, to transform these values into a single one representing potential, 22 different functions, which will not be described, were developed.

In Nature, when an active emotion is elicited the effect is not the same as if it were elicited for the first time [4]. This work uses, for all emotion types, function (2), based on [3], where $i$ is the intensity, $t$ is the emotion's threshold, and $p$ is the potential:

$$\text{reinforce(i, t, p)} = \log_2(\exp(i + t) + \exp(p)) \ . \tag{2}$$

The two global variables focused in this work are *arousal* and *mood*. Arousal is related to the physiological manifestation of emotions. It is characterized as follows: is positive; decays linearly with time; reinforcement occurs with emotion eliciting; increases elicited emotion potential. Mood refers to the longer-term effects of emotions. Moods can last for hours, days, and maybe longer, in contrast to emotions which last few minutes [4]. It is characterized as follows: can be negative or positive; converges to zero linearly with time; reinforcement occurs with emotion eliciting.

### 3.3    Dialogue's Emotional State

Suppose that, at a certain instant, a dialogue is the story's highest priority point of interest and, thus, is being focused by the expression channels. If each participating character is characterized by a different local emotional state, how is the global dialogue emotional state characterized? This work proposes a simple answer: *the dialogue's emotional state is the average of all the participant characters' emotional states*. In concrete, this corresponds to averaging each of the characters' active emotions intensities and global variables values.

## 4    Cinematography Expression

Cinematography environment expression is about telling a story through a camera. The section begins by describing some of the cinematography literature's established guidelines relating camera parameters to emotion expression and, then, proceeds to describe their application in this work.

### 4.1    Background

A *shot* represents a camera configuration of a certain time duration which is not broken up by cuts [7]. A shot can be either static or dynamic. Shots can vary, among others, according to the *distance* and to the *angle* with the point of interest. Regarding distance, the closer the camera is, the higher is the audience's attachment to the point of interest [5][7]. Five distance shots are commonly used [5]: (1) extreme close up, which focuses a particular detail, like the character's eyes; (2) close up, which focuses the character's face; (3) medium shot, which focuses the character from the waist up; (4) full shot, which focuses the whole character; (5) long shot, which films the whole character and also the surrounding environment. These shots need not focus characters, as any other point of interest can be focused as long as the distances are adjusted. Regarding angle, [5] mentions three representative shots: (1) *eye level* – the camera is placed at the height of the point of interest, representing a neutral view; (2) *high angle* – the camera films the point of interest from above creating the impression of smallness and isolation; (3) *low-angle* – the camera films the point of interest from below creating the impression of a powerful point of interest.

### 4.2    Expression

Cinematography expression reflects the focused point of interest emotional state's strongest emotion as follows:

(1) *If it is anger or pride*, a low-angle shot is chosen (Fig.2-a);

(2) *If it is fear*, a high-angle shot is chosen (Fig.2-b);

(3) *If its potential is on the interval [0; 1.5[*, the full shot is chosen (Fig.2-c);

(4) *If its potential is in the interval [1.5; 2.5[*, the medium shot is chosen (Fig.2-d);

(5) *If its potential is in the interval [2.5; 4.5[*, a close up is chosen (Fig.2-e);

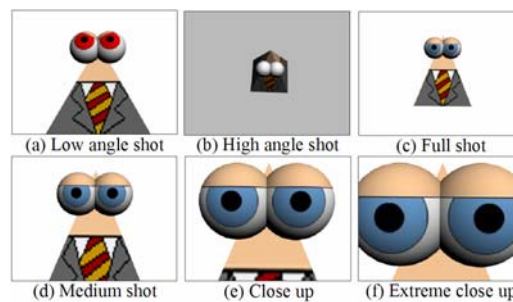(6) *Otherwise*, an extreme close-up of the eyes is chosen (Fig.2-f).



**Fig.2.** Cinematography expression shots reflect the character's emotional state.  a)-b) Angle shots reflect the character's power. c) –f) Distance shots reflect different emotion intensity

## 5    Illumination Expression

Illumination environment expression is about telling a story through lights. The section begins by describing research in illumination and on the relation between color and emotion proceeding, then, to describe its application to this work.

### 5.1    Background

Regarding placement, the *three-point-lighting* technique is widely used in movies to illuminate characters [7]. It is a configuration composed of the following light "roles": (1) *key light* – which is the main source of light focusing the character; (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 character from the background. Regarding light color, color association with emotion is widely documented (see [8] and associated references). For instance, red is normally associated with something exciting or aggressive; yellow with something cheerful; green with nature and, thus, relaxing; blue with quietness; green-yellow with vomit and, thus, displeasing; grey is neutral; among others. Regarding brightness, it is known that well illuminated scenes are happy and cheerful and poorly illuminated scenes are mysterious and sad. [7]

## 5.2    Expression

Illumination expression uses three-point-lighting to illuminate the focused point of interest. In particular, the key light is a point light placed between the point of interest and the camera. Emotion expression is achieved through key light's parameters manipulation. In concrete, its color is associated to the strongest emotion type according to Table 1. Presently, this work considers 12 out of the 22 OCC emotion types. Finally, brightness varies with the strongest emotion intensity and valence. Variation is implemented through the light's attenuation factor according to equation (3) if the emotion is positive and equation (4) if it is negative.

$$\text{Attenuation}_{positive} = \min(0.5, 1 - \text{emotionIntensity} / \text{maxEmotionIntensity}) . \qquad \textbf{(3)}$$

$$\text{Attenuation}_{negative} = \max(0.25, \text{emotionIntensity} / \text{maxEmotionIntensity}) . \qquad \textbf{(4)}$$

**Table 1.** Explored OCC emotion types to color mapping

| OCC Emotion type | Color (RGB) |
|---|---|
| anger, reproach | red (255, 0, 0) |
| disappointment, fears-confirmed | grey (200, 200, 200) |
| disliking | green-yellow (220, 255, 0) |
| distress | dark grey (153, 153, 153) |
| fear, relief, neutral | white (255, 255, 255) |
| hope, liking, satisfaction | bright yellow (255, 255, 200) |
| joy | yellow (255, 255, 0) |

## 6    Music Expression

Music environment expression is about telling a story through music. The section begins by describing research relating music and emotion and proceeding, then, to describe its application to this work.

### 6.1    Background

The relationship between music and emotion can be explored on four dimensions: (1) *Structural features* – which relates the music's structure with emotions; (2) *Performance features* – which refer to the influence of the *interpretation* of the music; (3) *Listener features* – which refer to the influence of the listener's attitudes and cultural influences; (4) *Contextual features* – which refer to aspects of the performance and/or listening situation. Regarding structural features, 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. There are many others parameters which lie beyond the scope of this work. Regarding performance features, [9] says that the expressive intention of the

performer is converted into various cues during the performance. Regarding listener features, they can consist of musical systems that are shared by a culture, inference dispositions based on personality, prior experience, musical talent and valenced memory associations. Finally, contextual features refer to aspects of the context under which the composition is performed and listened to. [9]

### 6.2    Expression

In this work, music expression 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 in the library, music was selected according to the following simple criteria: (1) Positive songs have fast tempo and, if they have lyrics it should be positively valenced; (2) Neutral songs have medium tempo; (3) Sad songs have slow tempo and, if they have lyrics, it should be negatively valenced. Regarding the association of lyrics emotional valence to the music's valence, if the performer tries to convey the music's mood through cues (subsection 6.1), then it is reasonable to expect that the lyric's mood propagates to the performance's structural features.

## 7    Study

A study was conducted to assess the influence and relevance of environment expression to the audience's perception of the story characters' emotional state, as well as the adequacy of this work's approach for each expression channel.

The study was based on an application called *dancing solids*. This is a cartoon-like application which tells stories about male and female geometric solids seducing each other through dance. In the end, if the female likes the male they'll simply marry.

The study was organized into four parts: (1) *Subject Profile* – where the subject's profile was assessed; (2) *Emotion Perception* - where the subject was presented with one of seven OCC emotion types – anger, disliking, distress, fear, joy, liking, reproach – or neutral emotion expression with varying configurations of two of the expression channels – cinematography and illumination. The subject was then asked to guess the expressed emotions from a set of options which was provided; (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 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. The subject was then asked which was the preferred version.

The study was presented to 50 students at Technical University of Lisbon. Average age was 23 years. Regarding emotion perception, collected data revealed that: perception of distress, joy, liking, neutral was highly accurate (above 75%) even without environment expression; illumination color expression increased accuracy particularly for anger (from 13% to 43%), disliking (from 13% to 20%) and reproach

(from 46% to 60%); the cinematography channel emotion type to camera shot mapping, in general, did not influence accuracy. Regarding music emotion valence, average subject classification matched predictions for 92% of the music. Regarding stories interpretation, 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%).

## 8    Conclusions

This work proposes an integrated architecture for storytelling capable of expression through three channels: cinematography, illumination and music. Stories are organized into prioritized points of interest which can be characters or dialogues. Emotion synthesis is based on the OCC emotion theory. Emotion expression is achieved through the expression channels. At each instant, the highest priority point of interest is focused. The cinematography channel reflects the point of interest's strongest emotion type and intensity. The illumination channel uses three-point-lighting to illuminate the point of interest and the key light's color and brightness vary according to the strongest emotion type and intensity respectively. The music channel expresses the valence of the point of interest's mood by playing music with the same valence. Music selection was based on tempo and lyrics emotional valence.

Evaluation of this work confirmed the relevance of environment expression for emotion perception in storytelling. Regarding the proposed approach, the study revealed that: the emotion type to camera shots mapping in the cinematography channel needs further tuning; illumination color association with the emotion types is effective; mood valenced music selection based on tempo and lyrics emotional valence is sufficient to produce satisfactory results; and, finally, people prefer a version of a story which is told with environment expression than one which does not.

## References

1. Solmer, A.: Manual de Teatro. Temas e Debates (2003)
2. Ortony, A., Clore, G., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press (1988)
3. Paiva, A., Aylett, R., Dias, J., Sobral, D., Louchar, S., Zoll, C., Raimundo, G., Rebelo, F., Otero, N.: VICTEC Deliverable 5.3.1: Final Prototype of Emphatic Synthetic Characters, Chapter 2 (2004)
4. Picard, R.: Affective Computing. The MIT Press (1997)
5. Arijon, D.: Grammar of Film Language. Silman-James Press (1976)
6. Hornung, A.: Autonomous Real-Time Camera Agents in Interactive Narratives and Games - MhD thesis at the Laboratory for Mixed Realities (2003)
7. Birn, J.: [digital] Lighting & Rendering. New Riders (2000)
8. Kaya, N.: Relationship between color and emotion: a study of college students in College Student Journal (2004)
9. Juslin, P., Sloboda, J.: Music and Emotion: theory and research. Oxford University Press (2001)