Users' Response to Affective Film Content: a Narrative Perspective

Luca Canini[1], Steve Gilroy[2], Marc Cavazza[2], Riccardo Leonardi[1] and Sergio Benini[1]
(1) DII-SCL, University of Brescia, Via Branze 38, 25123 Brescia, ITALY
(2) School of Computing, Teesside University, TS1 3BA Middlesbrough, UK
{firstname.lastname}@ing.unibs.it, {s.w.gilroy, m.o.cavazza}@tees.ac.uk

Abstract

In this paper, we take a human-centred view to the definition of the affective content of films. We investigate the relationship between users physiological response and multimedia features extracted from the movies, from the perspective of narrative evolution rather than by measuring average values. We found a certain dynamic correlation between arousal, derived from measures of Galvanic Skin Resistance during film viewing, and specific multimedia features in both sound and video domains. Dynamic physiological measurements were also consistent with post-experiment self-assessment by the subjects. These findings suggest that narrative aspects (including staging) are central to the understanding of video affective content, and that direct mapping of video features to emotional models taken from psychology may not capture these phenomena in a straightforward manner.

1. Introduction

There is a growing interest in the description of affective content of video [5] for content-based indexing and personalization. One of the major challenges for the successful description of affective video content is to develop frameworks that could properly relate media content to the users affective responses.

In this paper, we report research on the understanding of affective content for videos, which attempts at restoring a narrative perspective, *i.e.*, one in which emotional reactions are analysed throughout the dynamic presentation of film, and are matched to basic video features, without a priori assumptions about their relationship to audience reactions.

This work pursues two main objectives. Firstly, it aims at relating users emotional responses, captured through physiological signals, to affective video content, in a way which would be compatible in the long term with narrative principles. This approach, which investigates the global cycle of affective presentation and user response, contrasts with the direct attribution of emotional properties to video features based on the psychological literature [5]. The key aspect here is to operate from the perspective of film dynamics, which is best placed to capture narrative aspects. Another one is to extend previous work on the direct mapping of affective video feature parameters to emotional models using real-time physiological measurements.

2. Previous and Related Work

Hanjalic and Xu [6] pioneered the analysis of affective video content, through an approach based on the direct mapping of specific video features onto the Arousal and Pleasure dimensions of the Pleasure-arousal-dominance emotional model (PAD) [9]. They described motion intensity, cut density and sound energy as arousal primitives, defining an analytic time-dependent (using video frames for the time dimension) function for arousal aggregating these properties. This mapping is inspired from previous literature, but is not validated through physiological measurements, which would be the method of choice to assess a time-dependent model. Furthermore, the examples of arousal mapping given in [5] refer to live sports events (football matches videos), whose properties may not transfer entirely to the case of feature films, which have different editing and whose soundtrack is of a different nature (includes non-diegetic material, does not include audience reaction).

Xu et al. [19] have also described emotional clustering of films for different genres, using averaged values of arousal and valence, deduced from video parameters. One inherent limitation of this clustering approach may have been to use for the purpose of clustering a categorical description of target user emotions, with no clear indication that these would be elicited by the viewing of traditional film genres. Their system performed better for action and horror films than for drama or comedy, which they attribute to the prominence of specific features in these genres. This could also be analysed as a more efficient detection of arousal-related features, which tends to characterise these two genres, over valence-related ones, as detected with the defined video fea-

ture set (*e.g.*, brightness and colour energy as valence features).

De Kok [3] has extended some of this work by refining the modelling of colours, in an attempt to achieve a better mapping onto the valence dimension. Kang [7] has described the recognition of high-level affective events from low-level features using HMM, a method also used by Sun and Yu [16]. Soleymani et al. [15] have studied physiological responses to films, exploring a wide range of physiological signals and investigating correlations between users self reports and the affective dimensions accessible through physiological measurements. Their study emphasised individual differences in affective responses with an in-depth analysis of the correlation between dimensional variables and video features for each subject.

From a different perspective, films have been used in psychological research to induce emotional responses. Rottenberg and Gross [10] point at specific advantages of using films, such as enabling the study of emotion waveforms over time, and have shown strong coherence between skin conductance measures and user experience across a film sequence. Finally, Rottenberg et al. [11] provide recommendations on how to produce film clips (in terms of scenes from typical films, clip duration, etc.). Kreibig et al. [8] have also found skin conductance measures to be the most discriminant physiological signals to differentiate between fear and neutral as well as between sadness and neutral.

2.1. Filmic Emotional Theories

Emerging theories of filmic emotions [17] [13] should give some insight into the elicitation mechanisms that could inform the mapping between video features and emotional models. Tan [17] suggests that emotion is triggered by the perception of change, but mostly he emphasises the role of realism of the film environment in the elicitation of emotion.

Smith [14] attempted to relate emotions to the narrative structure of films. He described filmic emotions as less character-oriented or goal-oriented, giving a greater prominence to style. Smith sees emotions as preparatory states to gather information, and, more specifically, argues that moods generate expectations about particular emotional cues. The emotional loop according to Smith is made of multiple (and redundant) mood-inducing cues, which in return makes the viewer more prone to interpret further cues according to his/her current mood. Smiths conclusion that emotional associations provided by music, mise en scene elements, color, sound, and lighting are crucial to filmic emotions, should encourage us in our attempt to relate video features to physiological responses.

From the above discussion, it would appear that the dynamic nature of emotion should better be studied using real-time responses rather than average values (in favour of that are also anticipatory emotions, specific narrative effects, the necessity for emotions to build up, and the existence of emotional markers scattered throughout the action). Also, the most commonly used video features (e.g. motion intensity, sound energy) do not differentiate readily between narrative actions and their presentation. In addition, the distinction between diegetic and non-diegetic, or between narrative and presentation, may be less relevant than anticipated by appraisal-based models of emotions.

Finally, there are conflicting views on the extent to which emotional responses to films depend on the individual. Soleymani et al. [15] have studied individual differences on the physiological correlates of emotion. Tan [17] notes that, for traditional cinema, emotional responses are relatively uniform across different subjects, and Smith [14] notes that the same type of dependable emotions are generated across a range of audiences, despite individual variations. We shall thus not put individual variability at the heart of our experiments, and whenever we carry out measurements we will consider average values over our test subjects.

3. Experiments

Our experiments measured GSR as an indicator of arousal. Andreassi [1] relates that changes in both tonic and phasic measures of skin conductance are correlated with level of arousal. The skin conductance level (SCL) is the baseline tonic level at any moment in time, and the phasic skin conductance responses (SCRs) are momentary fluctuations from this baseline. We expected that a film that is more arousing should elicit a higher average SCL, and a greater number of SCRs with a greater magnitude. We also posit that the dynamic changes in arousal as indicated by GSR should correlate with features in filmic properties.

Experiments were conducted with 10 participants with a median age of 29. Data from two subjects was discarded following equipment dysfunction leading to partial loss of data, leaving us with data from 8 subjects (this number is however similar to other studies such as Soleymani et al. [15]). Subjects viewed a series of four film clips, with intervals of relaxing sound and visuals to calibrate a baseline SCL for each clip, and to mitigate the effect of the previous clip when viewing the next. GSR measurements were collected using a ProComp Infiniti™ data collection device with a skin conductance sensor, and recorded using the Biograph Inifiti™ software package. Skin conductance measurements were taken at a rate of 256 samples per second, and analysed via the smoothed output of the software at 8 samples per second. The sensor electrodes were placed on the second and third medial phalanges of the non-dominant hand, as recommended by Venables and Christie [18]. GSR measurements across participants vary in absolute levels of skin conductance, so have been normalized using baseline

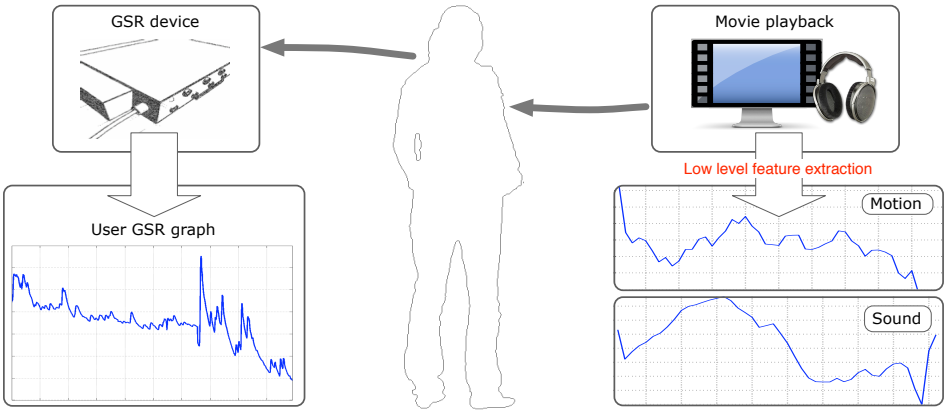

Figure 1. Overview of the setup for our experiments: the users galvanic skin response is recorded during clip viewing.

measurements. The overall setup is described on Figure 1.

Four short clips were used from three feature films: The Matrix, A Beautiful Mind, and Sweet & Lowdown. These films were chosen to show contrast in the filmic properties mentioned above, both in films of different genres, and between two scenes within the same film (A Beautiful Mind). The clips were between 1:38 and 2:11 minutes long, with a period of relaxation between each clip lasting 30 seconds. This was made up of a middle-grey screen with low-sound energy, down-tempo music. Clip duration was compatible with the low-end of the duration range described by Rottenberg et al. [11] necessary to induce an emotional response.

In our study, we consider average GSR responses over the 8 subjects, following the approach described in [11]. Subjects were also asked to give a subjective emotional evaluation of each film clip as a method of correlating GSR indications of arousal with subjective feelings of arousal towards each film. Questions were asked using 5-level Likert scales, and based on a Pleasure-Arousal model of emotion (such as Russells circumplex [12] or the Pleasure and Arousal dimensions of the PAD model [9]).

Physiological (GSR) responses will be plotted against video affective content. We consider motion and sound dynamics as the low-level video features to be correlated to the arousal response triggered by film viewing. In order to extract these video features, we first segment the video into shots (i.e. a sequence of continuous frames filmed with a single camera take, see [4] for a complete overview on the problem of shot detection). Then, we process each shot for the extraction of the related features, as proposed in [2]. We also use shot segmentation in the overall display of results (Figures 2 and 3), to support annotation with specific narrative events. In particular, the motion value is given by a combination of two elements: one related to the length of the shot, and the other given by the motion activity of objects and camera which captures the intuitive notion of action through the standard deviation of motion vector modules. The shorter and more motion intensive the shot, the higher the value. The audio track is divided according to the shot subdivision and for each shot the log-energy value is computed from a 8 kHz single-channel signal. To highlight the presence of brief and intense events (like thunders, screams etc), only audio samples whose modulus is above an adaptive threshold are taken into account. Audio and motion dynamics are normalised and summed together (with equal weights) to obtain a unique curve representative of the multimedia sequence. This curve has been smoothed with a low-pass filter to handle the persistence of affective features and it is the base to measure the correlation with the user arousal response.

4. Discussion

Figures 2 and 3 present the evolution of affective video features together with the average of normalized GSR response for the subjects. By using this physiological measure and this set of audiovisual features, we have deliberately chosen to explore arousal only. The rationale for that is that both its physiological correlates and its associated video features tend to be easier to describe and measure, from previous literature on the subject [11]. Our approach of averaging skin conductance signals over the subject sample and plotting it against time (or frame number) is compatible with the experimental setting described by [11], as is the comparison of the dynamics of the two curves. The time axis for both curves is segmented according to the various shots, and includes key narrative events depending on the film genre (for instance, action scenes in The Matrix).

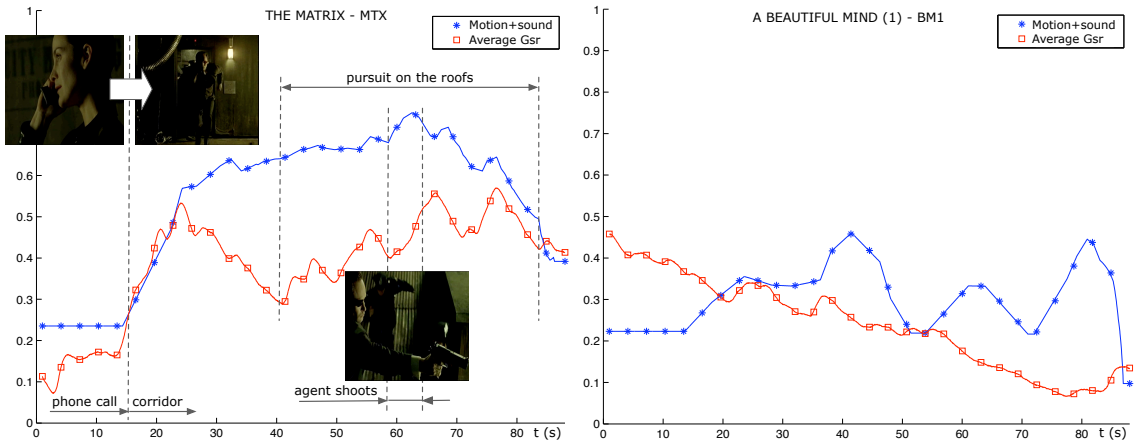

Figure 2. Experiment results (I): average normalized GSR for the subject population is plotted over time against affective video features. Dashed lines delimitate key narrative events. Films used in the experiment are 'The Matrix' (left) and 'A Beautiful Mind' (right).

4.1. Event-Based Analysis

The first level of analysis consists in examining the correlation between arousal values as measured by skin conductance and as expected from affective video features.

For the first video (The Matrix - MTX) we observe a strong correlation in arousal levels throughout the video duration (Figure 2, left). Even more significant is that variations in arousal are aligned for both user responses and affective video features. This is most visible for the surge at 15s, the subsequent increase between 40 and 60s, and the decreasing trend after 60s. In terms of correlation with narrative actions, the most significant phenomenon is the joint increase around 60s corresponding to the agents characters opening fire on Trinity. However, narrative coherence is observed throughout the experiment, as indicated by the fact that the patterns of variation as well as the arousal levels correspond to the specific phases identified, both on a shot basis and on the identification of narrative actions.

For the second video (A beautiful mind (1) - BM1) little correlation can be found between arousal levels for users response and for video features (Figure 2, right). However this takes place in a context where the overall arousal, based on video affective features, is rather low.

The third video (A beautiful mind (2) - BM2) shows moderate levels of arousal in terms of physiological response, despite the values of affective video features, for the first 90s of film viewing (Figure 3, left). In terms of response dynamics, a significant correlation can be found in terms of arousal response in the first 25s and around the key narrative transition at 95s.

The fourth video (Sweet & Lowdown - S&L) shows a moderate correlation during the onset, which soon evolves into a very good alignment between arousal indicators levels (Figure 3, right). Once again the dynamic response is the most significant, since arousal variations are co-occurrent to narrative action (at 20s). The different patterns of correlation follow once again the detected cuts, adding to the argument that narrative aspects may play a central role in the process.

4.2. Statistical Analysis

Further to the qualitative observations of the compared dynamics of physiological responses and affective video features, we have analysed statistical properties of the average values. In particular, we looked at the subjective scores of arousal given by the subjects, described in Section 3, compared to GSR. While the arousal scores were distinct for each clip, these need to correlate with the GSR measurements to be able to use GSR as a surrogate measure of arousal. Data were indeed positively correlated in all cases, as shown in Table 1 through the *correlation coefficient* (R) and the *coefficient of determination* (R^2), within a 95% confidence interval ($p < 0.05$).

Table 1. Correlation of Subjective Arousal Scores with Mean Normalised GSR.

	R	R^2	$p - value$
MTX	0.77	0.60	< 0.05
BM1	0.82	0.67	< 0.05
BM2	0.84	0.70	< 0.05
S&L	0.83	0.70	< 0.05

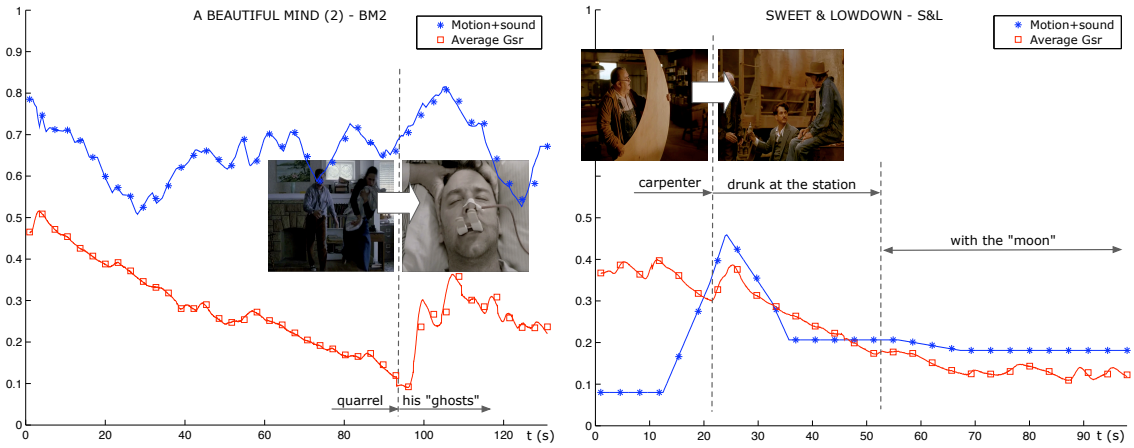

Figure 3. Experiment Results (II): films used in this experiment are 'A Beautiful Mind' (excerpt 2) and 'Sweet & Lowdown'.

The proposed correlation is between normalized GSR values and arousal score in an individual, and regression has been done for subject data points separately for each clip.

5. Conclusions

Whilst there have been several reports on the emotion-eliciting properties of films, and several proposals for affective video description using video features, no empirical study had attempted to relate the dynamics of users response to the real-time values of affective video features. In this study, we have established significant correlation between users physiological responses and affective video features for the arousal dimension. Following fundamental work on emotional responses to film [11], it appears that the affective dynamics are an essential component which may not be captured by measures of average values; nor could these be easily related to narrative aspects. Our results are compatible with previous findings, whilst still raising a number of questions deserving further investigation. Elements to be investigated will include cut density as well as a confirmation of Smiths hypothesis of emotional cues [13]. Secondly, it would appear necessary to further explore the various narrative determinants in terms of action and in terms of filming and editing. This could lead to a better understanding of the respective role of content and presentation in the elicitation of affective responses and could have an impact of various forms of video analysis and annotation.

6. Acknowledgements

This work has been funded in part by the European Commission under grant agreements CALLAS (FP-ICT-034800) and IRIS (FP7-ICT-231824).

References

[1] J. Andreassi. *Psychophysiology: Human Behavior and Physiological Response*. Routledge, 5th edition, 2007.

[2] L. Canini, S. Benini, P. Migliorati, and R. Leonardi. Emotional identity of movies. In *Proc. of the International Conference on Image Processing*, Cairo, Egypt, Nov. 2009.

[3] I. de Kok. A model for valence using a color component in affective video content analysis. In *The 4th Twente Student Conference on IT Proceedings*, Enschede, January 2006.

[4] A. Hanjalic. Shot-boundary detection: unraveled and resolved? *IEEE Transaction on Circuits and Systems for Video Technology*, 12(2):90–105, 2002.

[5] A. Hanjalic. Extracting moods from pictures and sounds. *IEEE Signal Processing Magazine*, 23(2):90–100, March 2006.

[6] A. Hanjalic and L.-Q. Xu. Affective video content representation and modeling. *IEEE Transactions on Multimedia*, 7(1):143–154, February 2005.

[7] H.-B. Kang. Affective content detection using HMMs. In *ACM international conference on Multimedia Proceedings*, Berkeley, CA, USA, November 2003.

[8] S. Kreibig, F. Wilhelm, W. Roth, and J. Gross. Cardiovascular, electrodermal, and respiratory response patterns to fear and sadness-inducing films. *Psychophysiology*, 44, 2007.

[9] A. Mehrabian. Pleasure-arousal-dominance: a general framework for describing and measuring individual differences in temperament. *Current Psychology: Developmental, Learning, Personality, Social*, 14:261–292, 1996.

[10] J. Rottenberg and J. Gross. *Emotion Elicitation Using Films*. Oxford University Press, New York, NY, 2004.

[11] J. Rottenberg, R. Ray, and J. Gross. Emotion elicitation using films. *In The Handbook of Emotion Elicitation and Assessment, J.A. Coan, J.J.B. Allen*, 2007.

[12] J. A. Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39:1161–1178, 1980.

[13] G. M. Smith. *Local Emotions, Global Moods, and Film Structure*, pages 146–167. In Passionate Views: Film, Cognition and Emotion, C. Plantinga and GM Smith, Eds. John Hopkins University Press, 1999.

[14] G. M. Smith. *Film Structure and the Emotion System*. Cambridge University Press, Cambridge, 2003.

[15] M. Soleymani, G. Chanel, J. Kierkels, and T. Pun. Affective ranking of movie scenes using physiological signals and content analysis. In *ACM workshop on Multimedia semantics Proceedings*, pages 32–39, Vancouver, Canada, October 2008.

[16] K. Sun and J. Yu. Video affective content representation and recognition using video affective tree and hidden markov models. In *ACII '07: Proceedings of the 2nd international conference on Affective Computing and Intelligent Interaction*, pages 594–605, Berlin, Heidelberg, 2007.

[17] E. Tan. Film-induced affect as a witness emotion. *Poetics*, 23(1):7–32, 1995.

[18] P. Venables and M. Christie. *Mechanism, instrumentation, recording techniques and quantification of responses*, pages 1–124. In Electrodermal Activity in Psychological Research, W.F. Prokasy and D.C. Raskin, Eds. Academic Press, New York, 1973.

[19] M. Xu, J. Jin, S. Luo, and L. Duan. Hierarchical movie affective content analysis based on arousal and valence features. In *ACM international conference on Multimedia Proceedings*, pages 677–680, Vancouver, Canada, 2008.