Presented at Intelligent Environments, Seattle, 21-22 July 2008

Towards end-user physiological profiling for video recommendation engines

Christos Calcanis
Chimera
University of Essex, UK
ccalca@essex.ac.uk

Vic Callaghan
Department of Computer Science
University of Essex, UK
vic@essex.ac.uk

Michael Gardner
Chimera
University of Essex, UK
mgardner@essex.ac.uk

Matthew Walker
Chief Technology Office
BT, UK
matthew.d.walker@bt.com

Keywords: affective computing, end-user profiling

Abstract

This paper describes research aimed at creating intelligent video recommendation engines for broadband media services in digital homes. The aim of our research is to harness physiological signals to characterise people's video selection preferences which we plan to integrate into new generations of video recommendation engines. We describe an initial experiment aimed at determining whether videos produce useable variations in physiology and linking these with emotional changes elicited by video material. We discuss our results and consider the possibility of utilising physiological sensing methods to build profiles that can be treated as signatures. Finally, we conclude by describing the future directions of our work.

1 Introduction

It has been proven that the "cognitive as well as the emotional state of a human is manifested in specific physiological reactions controlled by the autonomic nervous system" [12]. In this study we recorded physiological responses of human subjects to film material over time. By using affective computing techniques, we aim to show that there exists a very strong correlation between biophysical responses of different subjects with similar emotional responses to the same material. In that way, it becomes possible to create personalised signatures that encapsulate people's emotional response to film. By comparing these signatures, we can characterise people's video selection preferences, improve the performance of recommender systems and create methods for automated video analysis/highlight extraction. In the following sections we present an overview of affective computing techniques that utilise physiology. The experimental setup is described and results analysed. Finally, directions for future work are outlined.

2 Background

The most widely used techniques in affect recognition involve facial expressions and body gestures [3][17], speech [1][4], physiology [10][11] or a combination of multiple modalities.
Affect recognition methods based on physiological signals have been widely researched and offer some advantages when compared to alternative techniques. They are well-suited for realtime use [10] and can recognise a plethora of distinct emotion classes. The only requirement is the use of a biophysical sensor which is usually small and non-intrusive. This means that the subject can completely forget about the sensor and concentrate on the task or experiment at hand. The nature of the signals also implies that the system is immune to noise and that the subject is free to move around and conduct his daily activities, which is important when trying to operate in a realistic situation. Finally, biophysical signals can not easily be faked and provide a very accurate stream of information [2].
In [7], Lang proposes that all emotions can be described in terms of arousal (degree of excitement) and valence (negative/positive emotion). In [14] we learn that there exists a link between physiological signals and arousal/valence since *the activation of the automatic nervous system changes while emotions are elicited*. The way this link is manifested is with skin resistance reliably indicating arousal [13] and heart-rate indicating valence [5]. Many affective computing projects that utilise biophysical signals are focused on discrete classifications of emotion. The theoretical model that makes this possible is frequently that proposed by Russel in [15]. It is a two dimensional projection of common emotional states on the arousal and valence dimensions (See Figure 1 for some common emotions projected onto the map).

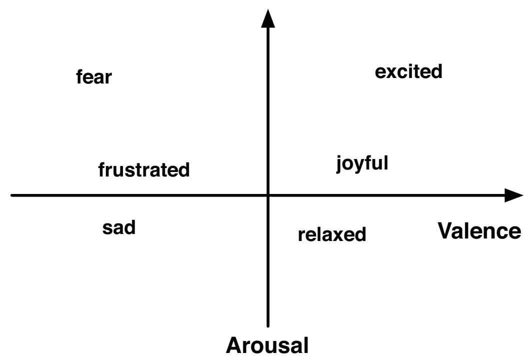

Figure 1: 2d Arousal-Valence Map

In practice, machine learning algorithms are used for learning and classification [11]. The human subjects are usually shown a combination of pictures (frequently from the International Affective Picture System [8]) and video material that is designed to elicit particular emotional responses. Biophysical measurements are taken and the subjects are asked to self-report on their emotional state. After a training phase, the classifier is ready to separate between the emotional states investigated. This approach has been successful but it is focused on recognition of discrete emotions and does not take into account the time dimension. Moreover, when video material is used as means to elicit emotion, emphasis is placed on short clips that are designed to elicit specific emotions. No attempts have been made to track correlations between highlights in video and biophysical responses and, in addition, profiles that model these correlations have not been built. Nevertheless, the success of this approach in detecting specific emotions from physiology provides us with the necessary foundation required to explore these relationships and track them over time.

In [16] the authors conducted an experiment to study the relationship between emotions and learning. As part of the experiment, the skin resistance of a human subject was recorded over time and taken as a measure of arousal. The experiment was composed of several learning sessions and one TV session where the subject watched video material while biophysical measurements were recorded. It was found that *when the participant was learning, the arousal was usually moderate and remained relatively stable but during the TV session the arousal varied greatly*. These results are encouraging as they seem to indicate that changes in arousal caused by a TV programme are manifested in the skin resistance measurements. This is not further researched since emphasis is placed on the relationship of learning to different emotional states.

Finally, in [6] and [5] the authors present a framework that automatically creates emotional profiles from video material. The profiles are based on the expected response of the video and are generated from video features such as sound intensity and shot length that are correlated with arousal and valence. One of the consequences of using features present in the video to extrapolate arousal and valence is that the resulting profiles are not personalised since they only capture expected responses of viewers. Nevertheless, Hanjalic et al discuss how they successfully used their framework for "highlight extraction" and propose its application to personalised recommendations.

3 Experiment

3.1 Sensor

The physiological sensor used for the experiment was the eXperimental Vital-sign-based Emotional State Transmitter (XVEST) as described in [10] and [16]. It is a biosensor that uses a fingerclip to sample blood volume pressure (BVP), heart rate (HR) and skin resistance (SR). Moreover, the sensor is small and not very intrusive. There was no discomfort or distraction reported by any of the subjects. Custom software was used to capture these signals in real-time at 14Hz sampling rate, timestamping and storing each reading.

3.2 Subjects and Video Material

Following a similar approach to [11], three twenty minute video clips extracted from full length hollywood feature films were used. The first is a highly intense comedy clip from the film *Four Rooms (1995)*. The second is an action sequence taken from *Heat (1995)*. Finally, the third is a highly suspenseful sequence from *28 Weeks Later (2007)*. Physiological measurements were recorded from seven subjects from culturally diverse backgrounds, all being postgraduate students at the University of Essex. Their age, sex and emotional state prior to the experiment are summarised in Table 1. Almost half of the subjects were stressed because University examinations were taking place at the time. The rest did not have similar sources of distress and described their emotional state as positive or normal.

Subject	*Sex*	*Age*	*Emotional State*
1	Female	25	Very Positive
2	Female	23	Very Positive
3	Male	25	Normal
4	Male	28	Extremely Stressed
5	Female	30	Normal
6	Female	25	Stressed
7	Male	27	Stressed

Table 1: Human Subjects

3.3 Procedure

The experimental procedure was the same for all subjects. Everyone was asked to wash their hands and were then connected to the physiological sensor and encouraged to relax and immerse themselves in the movie. An initial period of two minutes, defined as relaxation period, was used to collect physiological readings that were used for setting the norm. The clip was then started, and the researcher exited the room. On conclusion of the clip, the researcher entered the room, stopped signal capturing and conducted a short interview with the subject about the clip watched. The subjects explained their reaction to the video clip with an emphasis on discovering how the highlights in film matched periods of perceived emotional intensity.

3.4 Results

In this section, experimental results for skin resistance are presented. The average skin resistance recorded during the relaxation period is used to normalise all subsequent readings according to the formula used in [11]:

$$normalizedSR = \frac{rawSR - relaxSR}{relaxSR} \quad . \qquad (1)$$

Subjects 1 and 2 were in a very positive frame of mind before the experiment and found the comedy clip extremely amusing. This is reflected in the results as an overall rise or stable skin resistance with sharp drops to indicate periods of extreme amusement, laughter. Normalised SR graphs for Subjects 1 and 2 are presented in Figure 2 and Figure 3. Although the subjects come from completely different cultural backgrounds and have not previously seen the comedy clip, they display very similar responses to it. The clip builds up emotionally after the middle and the most emotionally active parts of the clip can be matched with the biophysical responses exhibited in an accurate way. Visually, this can be seen in Figures 2 and 3 as sudden drops in SR around samples 10000 and 15000 which correspond to the aforementioned emotionally stimulating parts of the clip.

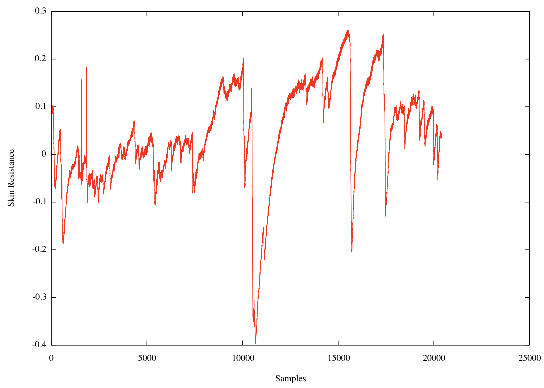

Figure 2: Normalised SR for subject-1 during comedy clip

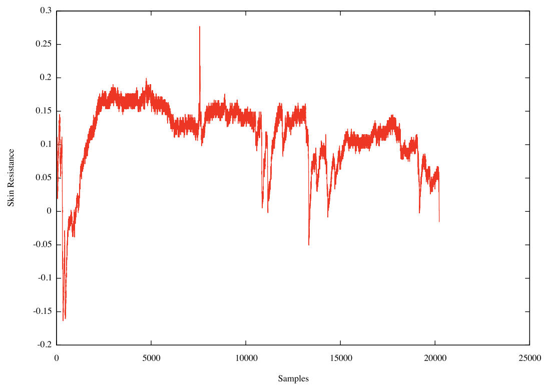

Figure 3: Normalised SR for subject-2 during comedy clip

Moreover, both subjects demonstrated extremely high SR values (above 1000) compared to the others. This might be explained by their relaxed and stress-free emotional state prior to the experiment. Raw SR values for the same clip are presented in Figure 4 and Figure 5.

Similarly, subjects 3 and 7 exhibited promising results for the action clip. The normalized SR graphs are presented in Figure 6 and Figure 7. Subject 3 was familiar with the clip as he had watched the movie many times in the past. On the other hand subject 7 was not familiar with the clip and found it exciting. The striking similarities in their biophysical response, even though the subjects come from completely different cultural backgrounds, seem to support the

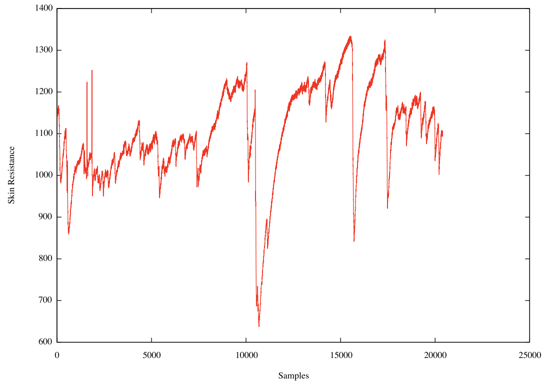

Figure 4: Raw SR for subject-1 during comedy clip

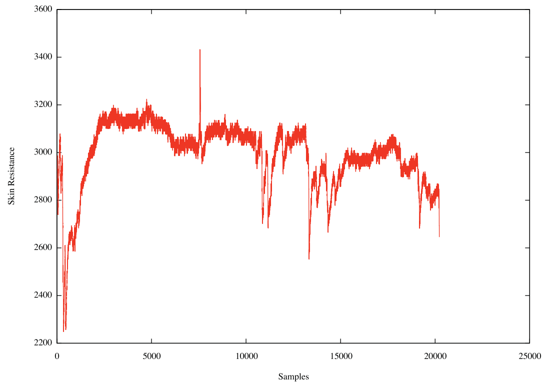

Figure 5: Raw SR for subject-2 during comedy clip

feasibility of emotional profiling over time. In detail, the two graphs are almost identical in the way the variations were exhibited. Sudden drops in skin resistance manifested themselves in the same place on the time axis and were of approximately the same duration. The only difference is the magnitude of the variations, with subject 3 showing diminished SR changes whilst subject 7 showing increased SR changes. This seems to be consistent with their previous experience of the clip, subject 3 having seen it before.

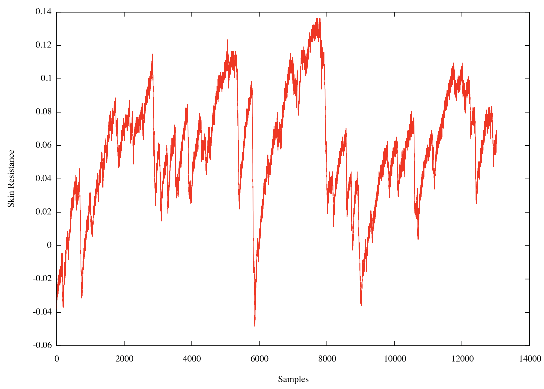

Figure 6: Normalised SR for subject-3 during action clip

The rest of the subjects showed no meaningful variations

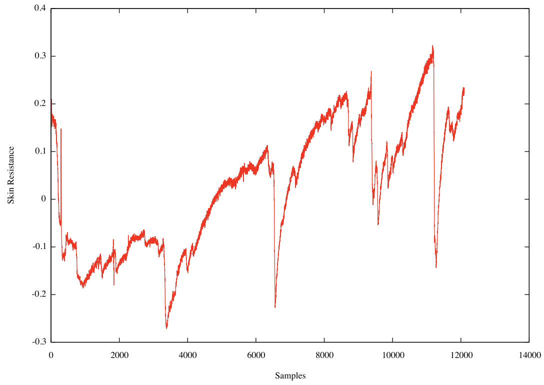

Figure 7: Normalised SR for subject-7 during action clip

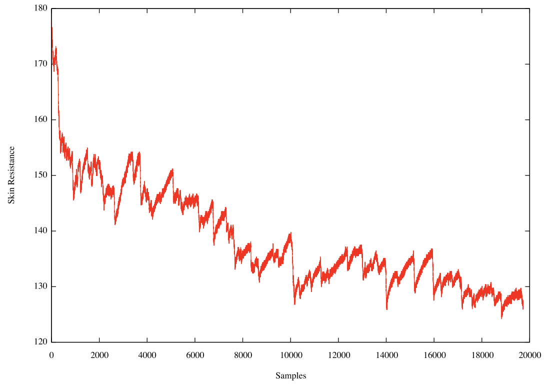

Figure 9: Raw SR for subject-5 during comedy clip

in the physiological data collected across all video clips. Most were under extreme stress and they could not relax and immerse themselves in the movies. Skin resistance was extremely low for all. Subject 5 provides us with a typical biophysical response seen in Figure 8 and Figure 9 that indicates complete disassociation from the video material. That particular subject also had problems with the English language spoken in the clip. Although subtitles were provided, it was very hard for her to quickly read them and keep track of the movie. The end result was that the clip had no emotional impact on her and this is reflected in the graphs.

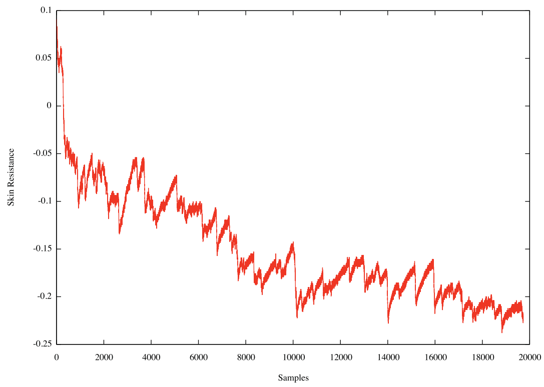

Figure 8: Normalised SR for subject-5 during comedy clip

4 Conclusions & Future Work

The results seem to support the hypothesis that physiological responses can be used to correlate the emotional responses of humans to video material. Whilst these are only preliminary findings and need more exhaustive investigation, we find them sufficiently heartening to encourage us to continue this research. In particular, the similarities in biophysical response (SR) between subjects who found the clips interesting and emotionally stimulating can be used as the basis of further research into profile building. In summary the results from our experiments indicate:

- Subjects that had similar emotional responses to the same material exhibited similar biophysical responses in terms of SR
- An overall stable/rising skin resistance could indicate a film that is relaxing or uplifting or amusing
- An overall falling skin resistance could indicate boredom or disassociation from projected material
- The emotional impact in terms of arousal of a video clip can be measured by the number and intensity of variations in skin resistance
- It is possible to accurately match intense variations in skin resistance with periods of intense emotional activity in video
- Normalisation of skin resistance with an average value obtained during a relaxation period is a promising approach towards building arousal profiles for different subjects

Although the experimental results are promising, they need to be validated with new experiments that utilise a larger set of subjects. In addition, more formally constructed samples, based on the Affect Intensity Measure (AIM) test [9], for determining a person's emotional state prior to the experiment will be integrated into the new experimental phase. Most importantly, consistency with biophysical responses for subjects that were stimulated by the same clip needs to be more thoroughly demonstrated and the importance of additional biophysical signals such as heart rate and temperature needs to be investigated.

The longer term aim is to integrate this mechanism with a video recommendation engine with the view of increasing the quality of recommendations, but this work will remain to be done until we have developed a reliable system for correlating physiological responses to emotional reaction to video. Clearly, this is an ambitious vision, with much work to be completed and we look forward to reporting out findings at the various stages of our research.

Acknowledgements

This work is supported by the Engineering and Physical Sciences Research Council and British Telecom. We would like to thank Chris Fowler for his valuable guidance and Malcolm Lear for biosensor technical design and support.

References

[1] N. Amir and S. Ron. Toward automatic classification of emotions in speech. *Proceedings ICSLP*, pages 555–558, 1998.

[2] G. Chanel, J. Kronegg, D. Grandjean, and T. Pun. Emotion Assessment: Arousal Evaluation Using EEG's and Peripheral Physiological Signals. In *MRCS*, volume 4105 of *Lecture Notes in Computer Science*, pages 530–537. Springer, 2006.

[3] L. S. Chen and T. S. Huang. Emotional expressions in audiovisual human computer interaction. *Proceedings ICME*, pages 423–426, 2000.

[4] F. Dellaert, T. Polzin, and A. Waibel. Recognizing emotion in speech. *Proceedings ICSLP*, pages 1970–1973, 1996.

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

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

[7] P. J. Lang. The emotion probe: Studies of motivation and attention. *American Psychologist*, 50(5):372–385, 1995.

[8] P. J. Lang, M. M. Bradley, and B. N. Cuthbert. International Affective Picture System (IAPS): Digitized photographs, instruction manual and affective ratings. *Technical Report A-6, University of Florida*, 2005.

[9] R. J. Larsen, E. Diener, and R. A. Emmons. Affect Intensity and Reactions to Daily Life Events. *Personality and Social Psychology*, 51(4):803–814, 1986.

[10] E. Leon, G. Clarke, V. Callaghan, and F. Sepulveda. A user-independent real-time emotion recognition system for software agents in domestic environments. *Eng. Appl. Artificial Intelligence*, 20(3):337–345, 2007.

[11] C. L. Lisetti and F. Nasoz. Using Noninvasive Wearable Computers to Recognize Human Emotions from Physiological Signals. *EURASIP Journal on Applied Signal Processing*, 2004(11):1672–1687, 2004.

[12] C. Peter, J. Waterworth, E. Waterworth, and J. Voskamp. Sensing Mood to Counteract Dementia. In *1st International Workshop on Pervasive Technologies for the support of Alzheimer's Disease and Related Disorders*, 2007.

[13] R. W. Picard and J. Scheirer. The Galvactivator: a glove that senses and communicates skin conductivity. In *9th International Conference on Human-Computer Interaction*, New Orleans (US), August 2001.

[14] H. Prendinger, H. Dohi, H. Wang, S. Mayer, and M. Ishizuka. Empathic embodied interfaces: Addressing users affective state. *Proceedings Tutorial and Research Workshop on Affective Dialogue Systems*, pages 53–64, 2004.

[15] J. A. Russel. A Circumplex Model Of Affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

[16] L. Shen, E. Leon, V. Callaghan, and R. Shen. Exploratory Research on an Affective e-Learning Model. In *Workshop on Blended Learning, International Conference on Web-based learning*, Edinburgh (UK), August 2007.

[17] L. C. De Silva, T. Miyasato, and R. Nakatsu. Facial emotion recognition using multimodal information. *Proceedings ICSP*, pages 397–401, 1997.