Towards detection of interest during movie scenes

Joep JM Kierkels
Computer Science Department University of Geneva
Battelle Building A, 7 Route de Drize CH - 1227
Carouge, Geneva, Switzerland +41(22) 379 0152
Joep.Kierkels@unige.ch

Thierry Pun
Computer Science Department University of Geneva
Battelle Building A, 7 Route de Drize CH - 1227
Carouge, Geneva, Switzerland +41(22) 379 0223
Thierry.Pun@unige.ch

ABSTRACT
In this paper a basic approach for the continuous detection of interest based on physiological signals is described and tested on data recorded during movie watching. Feature extraction, selection, and classification are used and an indicator f [t] for interest level is proposed. Using this indicator, interest changes can be detected, even within short movie scenes. No correlations are found between interest and the commonly used emotional parameters arousal and valence.

Categories and Subject Descriptors
J.4 **[Computer applications]**: Social & Behavioral – *Psychology*

General Terms
Measurement, Human Factors.

Keywords
Emotions, Interest, Biosignals, Signal Processing.

1. INTRODUCTION
When people are faced with something unknown, remarkable, or interesting, they usually try to pay close attention to it. They will be silent while observing, sit very still, and focus their gaze towards the object of interest. In an increasing number of areas, such as the advertisement business and the gaming- and movie industries, it is of great importance to know whether or not a newly designed product is able to attract and maintain the interest of an audience. In a user controlled environment, one can imagine a product, e.g. a game, adapting to the personal interest levels. To obtain such knowledge, one should be able to monitor how people respond to the product.

Currently, this is mainly done by e.g., using questionnaires or beta-testers, and by observing behavioural traits such as facial expressions and gaze direction [11]. Questionnaires depend on getting either verbal or written feedback from people, which may cause interruptions while focusing on a product, on a movie, or on music [4]. Such interruptions can affect the level of attention and interest that they are supposed to reflect and interruptions are therefore usually postponed. When a session has ended, only a single feedback measure is taken to describe the participant's level of interest. Clearly, this results in a suboptimal representation of the level of interest since this single value cannot reflect subtle changes throughout the session and may be biased towards the level of interest at the end of a session. Another way to observe a person's responses is to monitor the physiological responses of that person. Although theorists are still debating on the exact relation between emotions and physiological responses, it is undisputed that such a relation exists [3]. Like other emotions, interest is related to changes in human physiology and changes in several physiological features like heart rate, skin conductivity, and respiratory frequency can provide information on a person's level of interest. Using physiological responses for monitoring user interest levels has the key advantages that it does not require interrupting a user, it can be performed continuously, and it provides instantaneous unbiased measures. Because of these advantages, more and more studies on emotion assessment use recordings of physiological responses [2;7;13]. For detecting interest, only few studies have so far used physiological signals, focussing mainly on skin conductivity and ECG [1;9]. Apart from changes in physiological responses, emotions and thus interest also affect other modalities like facial features [10]. Examples of changes in these features are sometimes obvious, the corners of the mouth move upwards when one feels happy and the eyebrows move downwards when angry. Similar changes can be observed for interest as is illustrated in the first lines of this introduction. The approach discussed in this paper can also be applied to interest estimation based on facial features, or based on both facial features and physiological features. The difficulty of how to represent or display an emotion is to be addressed whenever it is required to self-assess emotions or when an estimated emotion needs to be visualised. To quantify and represent emotions, emotion theorists have developed a range of models. These include listings of 'basic emotions', tree structures of emotions, and emotion spaces such as the valence-arousal space [12]. In the valence-arousal space emotions are weighted on two separate scales. The valence scale indicates to what extent an event is pleasant, the arousal scale indicates the excitement-level that comes with the emotion, ranging from calm to excited.

In this paper, a method for the monitoring of interest, based on physiological responses, is proposed and is applied to the analysis of movie scenes. Multiple features are derived from physiological signals recorded during the viewing of movie scenes. The experiments described here were performed for a different study [Acknowledgment] and were already available. No improvements could be made to the protocol without going through lengthy new experiments. In a movie scene, many changes can occur and often it is the intention of a movie director to fluctuate the level of interest that a viewer experiences. On the one hand, this makes movie scenes highly challenging for

detecting interest levels, but on the other hand it is difficult to attribute observed changes to changes in interest. Other factors like screen brightness and sound volume are may also have effects on physiological signals.

From the physiological data that were recorded during the experiment, two classes of data with distinctive levels of interest are extracted. A Linear-Discriminant-Analysis (LDA) classifier is trained on these two classes. Subsequently the data recorded during the experiments are analyzed by the classifier. This results in a series of estimates that state at each moment in time the class to which the observed physiological signals most likely belong. It is worth noting that although all signal analysis and classification is performed offline, there are no conceptual obstructions to using online, near real-time, analysis. The resulting classifications are further used to address four research questions: 1) Can analysis of physiological responses alone be used to determine whether or not a person is viewing a movie? 2) Is there a correlation between the estimated level of interest and the genre (horror, drama, action, comedy) of a movie? 3) Is there a correlation between the estimated level of interest and the self-assessed emotion characteristics of arousal and valence? 4) Can changes in interest be observed during the movie scenes, separating between fragments within scenes?

By comparing the physiological data that are recorded whilst the participants are self-assessing their emotions, to data recorded whilst participants are viewing the movie scene, it is found that physiological data reveals whether or not a movie is being watched. From analysis of multi-participant data on single movie scenes it is possible to detect changes in interest within a single movie scene. Correlations between interest level and movie genre, arousal, or valence are not found.

2. PROTOCOL AND METHODS

Before starting the experiment, the participants were informed about the video contents. They then had a brief training about the self-assessment procedure and concerning the meaning of arousal and valence. A total of 7 participants, all volunteers, 4 male, age 30 (S.D.=6.6), participated in the experiment.

During the experiment, 64 scenes from eight movies were shown to the participant. The extracted scenes, eight for each movie, contained an emotional event (judged by the experimenter). The selected movies were from recent famous movies and mostly similar to those used in other studies [6]. Four major genres were represented in these movies: drama, horror, action, and comedy. The titles of the movies were: Saving Private Ryan (action), Kill Bill, Vol. 1 (action), Hotel Rwanda (drama), The Pianist (drama), Mr. Bean's Holiday (comedy), Love Actually (comedy), The Ring, Japanese version (horror) and 28 Days Later (horror).

Prior to each movie scene, a short neutral clip (approximately 30s) was shown. This was intended to normalise the emotional state at the start of each movie scene. After watching a movie scene, the participant filled in the self-assessment form on arousal and valence which popped up. The 64 movie scenes were displayed in random order. Neutral clips were selected from clips provided by the Stanford psychophysiology laboratory (http://www-psych.stanford.edu/~psyphy/resources.htm).
Average durations of the movie scenes and neutral clips were ~ 120 s and 30 s. During the experiment a participant was seated in a sound-isolated Faraday room while physiological signals were recorded. Respiration, electro-cardiogram (ECG), galvanic skin resistance (GSR), skin temperature, and blinking parameters from the electro-oculogram (EOG) were monitored using the BioSemi Active-two system (http://www.biosemi.com/). All signals were sampled at 1024 Hz and offline filtered using different combinations of DC correction, low-pass (LP), high-pass (HP), and band-pass (BP) filters, as indicated in Table 1.

Table 1. List of physiological signals with their filter-settings and the features derived from them

Signal		Feature	
1	Respiration (DC + 20 Hz LP)	1	Respiration Depth (arb.)
		2	Respiration Rate (s^{-1})
2	ECG (0.5-35 Hz BP)	3	Heart Rate (s^{-1})
		4	Heart Rate derivative (s^{-2})
		5	Inter-Beat- Interval (s)
3	GSR (DC + 3 Hz LP)	6	GSR (arb.)
		7	Number of peaks
		8	Peak amplitude (arb.)
		9	dGSR/dt (arb.)
4	Temperature (1 Hz LP)	10	Temperature (^oC)
		11	dTemp/dt ($^oC/s$)
5	Blinks (0.3-30 Hz BP)	12	Blink Amplitude (arb.)
		13	Blink Duration (s)
		14	Blink Frequency (s^{-1})

Table 1 also indicates which features are derived from the physiological signals. These features are frequently used for estimation and classification of emotions, and more information on these features and their derivation can be found in e.g.,[5]. For peak-detection-based features, like heart rate and respiration depth, a new heart rate is computed whenever a new peak is detected based solely on the distance between the last two peaks. Each of the features, $s_n[t]$, with $n = [1,2,\ldots,N]$ indicating feature number, is sampled at 128 Hz and is either obtained via down-sampling of a physiological signal or interpolation between events detected from the physiological signal. Subsequently, a 10 s moving average filter is applied to all features to enable feature comparisons between participants. The time index, t, is thus discrete and is bounded by the total duration, T, of each participant's recording session.

2.1 Feature selection

The next step is to determine which of the features are sensitive to changes in level of interest. To this end, each of the 14 features, $N = 14$, is separately analyzed. In order to detect changes in feature value that are affected by the ending of a scene, the 64 moments during the experiment which correspond to the ending of a movie scene are marked. Fragments of all features are taken from 10 s prior to and 10 s after these marked points, resulting in fragments that are aligned to the endings of movie scenes. Subsequently, the first 32 of these fragments are inspected. Clearly, when the movie scene ends the interest level of a participant will change since attention will shift from watching the scene to relaxation and filling in the self-assessment. For all 32 segments prior to scene ending, feature variance is computed and compared against the feature variance after scene ending, using a pair-wise t-test (significance level 0.95). If a significant difference between variances of a feature is found for more than 2 participants, this feature is selected as relevant for interest detection.

2.2 Classification

Using the relevant features, a two-class LDA classifier is trained on one participant. The training feature-vector of the 1st class contains all relevant features recorded 6 to 0 s before scene ending and the training feature-vector of the 2nd class contains all relevant features, recorded 1 to 7 s after scene ending. After training, the classifier is re-applied to the training data to check the accuracy of classification. Next, data of the whole experiment are classified.

3. RESULTS

The pair-wise t-test reveals that most of the features are relevant for detection of interest. Only duration of blinks, number and height of the GSR peaks, and the two temperature features have non-significant results. In further analysis, the heart rate feature is also excluded as it is inversely proportional to the inter-beat-interval feature. Classifier accuracy on the training data was close to 72%. Correct classification for elements of the 2nd class was higher (84%) than for those of the 1st class (60%).

Following these results, the classifier was used to classify the data recorded throughout the whole experiment. The resulting classifications are represented as $c[t]$, with $c[t] \in \{0,1\}$ for class 2 and class 1 respectively. In Fig. 1, results of this classification on a 600 s part of the experimental data are shown.

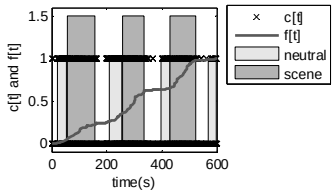

Figure 1. Example of classification results *c*[*t*], with cumulative function *f* [*t*], and synchronised with the protocol.

In Fig. 1, the class 1 labels appear to be grouped, which becomes even more obvious if a cumulative function $f[t]$, defined as

$$f[t] = \left(\sum_{0}^{t} c[t]\right) / \left(\sum_{0}^{T} c[t]\right),$$

is computed and displayed, as can also be seen in Fig 1. It should be noted that the groups of class 1 labels coincide with the playing of movie scenes, as can be seen from Fig 1. In this plot, the time intervals during which movie scenes or neutral clips were played are highlighted using the annotations to the data.

During the playing of movie scenes, $f[t]$ increases more rapidly than during self-evaluation, which could indicate that these events involve different levels of interest. In order to address the 2nd and the 3rd research questions as stated in the introduction, and thus to see whether interest level depends on movie genre, or self-assessed arousal and valence, a measure should be derived for the average interest level during each movie scene. For this, it is proposed to use the average slope of $f[t]$ over the duration of the specific movie scene. For all participants the average slopes of all movie scenes are computed and results for identical movie scenes are averaged over participants. Averaged slopes and standard deviations per movie scene are shown in Fig. 2A. Visual comparison between different movie genres, also indicated in Fig. 2A, shows no relation between interest level and movie genre. Self-reported arousal and valence values were averaged and are shown in Fig.2B.

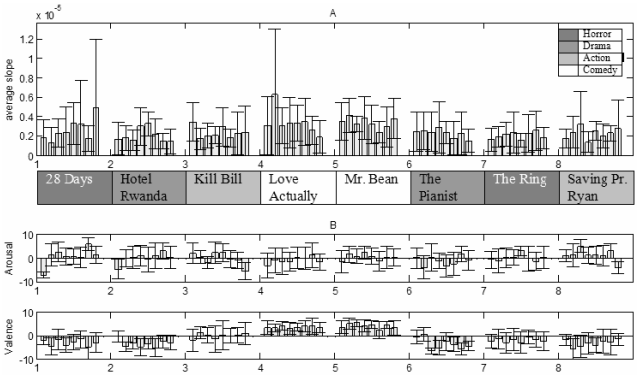

Figure 2. Average interest levels (A) and Average arousal and valence levels (B) per movie scene indicating movie genre.

Correlation coefficients, ρ, between all possible combinations of interest, arousal and valence were computed over all participants and all movie scenes. Results, shown in Table 2, reveal no significant correlation of interest with any of the other parameters.

Table 2. Correlations (ρ) between Interest, Arousal, Valence

Combination	ρ	Combination	ρ
Arousal-Valence	-0.04	Valence- Interest	0.19
Arousal-Interest	0.04		

To analyze whether changes in interest can be observed within single movie scenes, interest levels as indicated by $f[t]$ are again compared between participants. For finding the interest changes within the scene, the interval of $f[t]$ which coincides with a specific movie scene is normalised. Normalization implies that the minimum and maximum of $f[t]$ over this scene are scaled to 0 and 1 respectively. The normalised $f[t]$ is then averaged over all participants, as is illustrated for two movie scenes in Fig. 3.

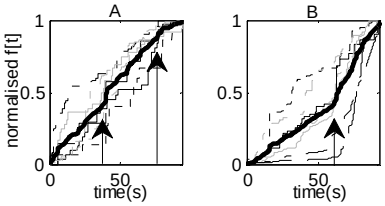

Figure 3. Normalised *f* [*t*] on two different movie scenes, illustrating the changes in interest level. Thin lines indicate single participants, thick line indicates average.

In the scene of Fig. 3A, taken from the movie "28 Days Later", the main character is inside a church with many people lying on the floor as if they were dead. After approximately 40 s, two people suddenly awaken. Following this there is a noise and a priest appears who tries to attack the main character. Approximately 80 s after scene onset the main character hits the priest and runs away. The scene of Fig. 3B, taken from the movie "Love Actually", starts near a lake where the main character is writing a book. After 5 s, pages of this book are blown into the water and two people jump in the lake to recover these pages. About 60 s after onset, the conversation between the two becomes humorous.

4. DISCUSSION

A classification accuracy of 72 %, on 2 classes (50% chance level), is moderate compared to state of the art accuracies reported

for detection of emotions, e.g., 84 % in [8]. However, our result does not require averaging over prolonged time periods. Adding extra, relevant features, e.g., systolic and diastolic blood pressure, finger pulse transit time, and electro-encephalography-based features, will probably increase accuracy. The proposed classifier fuses the modalities and features in an attempt to isolate the effects of interest. A two-class classifier was used rather than defining interest levels over a more continuous range, e.g., using regression, because our training data does not allow for validation of continuous scales. It should be noted that the data for training the classifier only reflect the ending of a movie scene and that it is hypothesised that changes in interest coincide with this. Obviously, other changes like changes in sound volume and screen luminance will also coincide with scene endings and they can affect classification if they induce changes in the same physiological signals. With the current protocol it is impossible to exclude the influence of these other changes as this would require setting up different experiments in which changes in interest level are isolated from other changes [10]. The current study provides sufficient indications that such a follow-up study with new experiments could be successful. For example, the results on detecting interest changes inside scenes, as illustrated in Fig. 3B, do not only reflect changes in luminance and sound but are also affected by changes in story line. The time points which were mentioned in text below the figure can clearly be seen in Fig. 1.

No resemblance was found between interest level and movie genre. This indicates that the preferences for movies are not solely determined by movie genre. Moreover, the overall genre of the movie may not have been reflected in each of the scenes taken from that movie. A scene taken from a drama movie can be consider e.g., funny or romantic, especially when isolated from the rest of the movie. Correlations of interest with valence and arousal were also not significantly high. This indicates that not just this single aspect of the experiencing of emotions is detected. Only a small correlation with valence was found which could reflect that scenes of positive valence are considered to be slightly more interesting than scenes of negative valence. Interpretation of the results obtained for interest within scenes requires annotations of the events in the movie scenes. These annotations should be as objective as possible and could in the future be based on the evaluations of many viewers. For now, the annotations are provided by the authors. The two scenes and their annotations given here are intended to illustrate the co-occurrence of changes in story line and $f\,[t]$. Such co-occurrences were found in 43 of the 64 scenes.

5. CONCLUSIONS

A method for the monitoring interest based on physiological responses is derived and applied to the analysis of movie scenes. By using features derived from standard physiological signals and using linear classifiers it was first shown that it is possible to classify with 72% accuracy whether a recording is taken prior to or after scene ending in the training set. The use of extra features or other classification methods, possibly taking into account temporal aspects, could further increase this accuracy. Then, considering that the ending of a scene coincides with a change in interest level, other research questions are addressed and it is found that physiological features do indicate whether or not a person is viewing a movie. Moreover, the physiological features can also be used to detect which parts within a movie scene are of particular interest to the (average) viewer. As this enables the detection of changes in interest level of an individual as well as the detection of averaged interest over groups of individuals it can potentially be of great value in other areas as well, like advertising or gaming. No strong correlations were found between interest and genre, arousal and valence.

This study has demonstrated the possibility to detect interest-related changes based on physiological signals and calls for other studies and experiments which target more purely on classifying between different levels of interest, like those in [10].

6. ACKNOWLEDGMENTS

The authors thank Mr. M. Soleymani for sharing data. His work using these data will be presented at the Machine Learning and Multimodal Interaction workshop 2008, Utrecht, the Netherlands.

7. REFERENCES

[1] Bolls PD, Lang A, and Potter RF, "The effects of message valence and listener arousal on attention, memory, and facial muscular responses to radio advertisements," *Communication Research*, vol. 28, no. 5, pp. 627-651, 2001.

[2] Chanel G, Kronegg J, Grandjean D, and Pun T, "Emotion assessment: Arousal evaluation using EEG's and peripheral physiological signals," *Proc. Int. Workshop MRCS, Special Session: Multimodal Signal Processing*, 2006.

[3] Cornelius R, "Theoretical approaches to emotion," *Proc ISCA workshop on Speech and Emotion*, pp. 3-11, 2000

[4] Gu R, Zhu M, Zhao L, and Zhang N, "Interest mining in virtual learning environments," *Online Information Review*, vol. 32, no. 2, pp. 133-146, 2007.

[5] Haag A, Goronzy S, Schaich P, and Williams J, "Emotion Recognition Using Bio-sensors: First Steps towards an Automatic System," in *Affective Dialogue Systems* Springer Berlin / Heidelberg, 2004, pp. 36-48.

[6] Hanjalic A and Xu LQ, "Affective video content representation and modeling," *IEEE Transactions on Multimedia*, vol. 7, no. 1, pp. 143-154, 2005.

[7] Lisetti CL and Nasoz F, "Using noninvasive wearable computers to recognize human emotions from physiological signals," *Eurasip Journal on Applied Signal Processing*, vol. 2004, no. 11, pp. 1672-1687, 2004.

[8] Picard RW, Vyzas E, and Healey J, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175-1191, 2001.

[9] Ravaja N, "Contributions of psychophysiology to media research: Review and recommendations," *Media Psychology*, vol. 6, no. 2, pp. 193-235, 2004.

[10] Reeve J, "The Face of Interest," *Motivation and Emotion*, vol. 17, no. 4, pp. 353-375, 1993.

[11] Richins ML, "Measuring emotions in the consumption experience," *Journal of Consumer Research*, vol. 24, no. 2, pp. 127-146, 1997.

[12] Russell JA and Mehrabian A, "Evidence for A 3-Factor Theory of Emotions," *Journal of Research in Personality*, vol. 11, no. 3, pp. 273-294, 1977.

[13] Sammler D, Grigutsch M, Fritz T, and Koelsch S, "Music and emotion: Electrophysiological correlates of the processing of pleasant and unpleasant music," *Psychophysiology*, vol. 44, no. 2, pp. 293-304, 2007.