Biometric Responses to Music-Rich Segments in Films: The CDVPlex

Alan F. Smeaton and Sandra Rothwell

CLARITY: Centre for Sensor Web Technologies
Dublin City University, Glasnevin,
Dublin 9, Ireland
Alan.Smeaton@DCU.ie

Abstract

Summarising or generating trailers for films or movies involves finding the highlights within those films, those segments where we become most afraid, happy, sad, annoyed, excited, etc. In this paper we explore three questions related to automatic detection of film highlights by measuring the physiological responses of viewers of those films. Firstly, whether emotional highlights can be detected through viewer biometrics, secondly whether individuals watching a film in a group experience similar emotional reactions as others in the group and thirdly whether the presence of music in a film correlates with the occurrence of emotional highlights. We analyse the results of an experiment known as the CDVPlex, where we monitored and recorded physiological reactions from people as they viewed films in a controlled cinema-like environment. A selection of films were manually annotated for the locations of their emotive contents. We then studied the physiological peaks identified among participants while viewing the same film and how these correlated with emotion tags and with music. We conclude that these are highly correlated and that music-rich segments of a film do act as a catalyst in stimulating viewer response, though we don't know what exact emotions the viewers were experiencing. The results of this work could impact the way in which we index movie content on PVRs for example, paying special significance to movie segments which are most likely to be highlights.

1. Introduction — Music and the Movies

Hearing music affects us emotionally, and most of us have employed music for just that effect at one stage or another. This can vary from soft music to help us unwind and relax, or something loud and rhythmic as we go about some housework ! Music can evoke intense feelings, which can sometimes manifest themselves in strong physical reactions or behaviour such as shouting or crying. A large-scale survey revealed that 8% of all crying episodes in the US were evoked either directly or indirectly by music [5]. Less extreme physical reactions are more common however, for example "goose-bumps" and lumps in the throat.

Music is an important component of film and historically has been used to provide a backdrop before the arrival of "talkies". Music affects how the audience will perceive a scene in a film — we see an actor shout, but it is the music which tells us that he is joyful, not angry. Like lighting, music sets moods and tonalities in a film, but it differs from other filmic elements in that we hear it as opposed to seeing it. However, music's influence can be felt on many levels, from its purely functional roles in serving as a "cohesive force" [12] and storytelling device, to its more elusive role in creating the atmosphere and colouring the tone of a film. We know that music has a significant effect on how people perceive emotion in general, and we know that film composers and music editors exploit this in how they use music in feature films.

It is widely accepted that there is a correlation between our emotional responses and our physiological functioning. Common physiological responses to music are activations of the sympathetic nervous system such as changes in heart rate, blood pressure, breathing, and skin temperature, and the release of arousal hormones from the adrenal glands. Many of these reactions can be easily and unobtrusively monitored.

The work reported here seeks to investigate three questions related to film, music and emotions. The first is whether emotional responses experienced by groups of people watching films can be detected through physiological measurement. Secondly, we ask whether individual viewers in a group experience similar physiological

reactions to emotional stimuli and thirdly whether the occurrence of music in feature films correlates with viewers' emotional responses. In the next section we give background details on our experimental setup and how we gathered biometric readings as groups of people watched a selection of films, and in section 3 we describe how we annotated a selection of films for their emotive content. Section 4 presents our results, correlating film-induced emotions with observations from viewings. Section 5 presents a brief summary of some related work and our conclusions are presented in section 6.

2. The CDV*Plex* — An Instrumented Cinema

In order to carry out experiments on film-evoked emotions, we re-created an environment that was as close to a true cinematic experience as possible. The CDV*Plex* is an air-conditioned windowless room in which a Dolby 5.1 surround sound speaker system, a DVD player and a large-screen digital projector were installed for showing films to small groups. Each chair was outfitted as a Smart Chair (described below). We chose a number of biometric measurement tools as described below.

Heart rate is a very reactive metric and responds quickly to stimuli. Fluctuations of greater than ± 10 b.p.m. from a baseline are interpreted as a response and anomalies such as peaks and plateaux can indicate that the subject experienced a change in emotional state. Large spikes in the heart rate are an indication that the subject may have undergone a surprise or shock. Similarly, a sustained period of increased heart rate such as a rise of 25% over the baseline maintained for 3 to 4 minutes are an indication of a period of continuous excitement. We used a Polar S610i Heart-Rate Monitor (HRM) for each participant in our experiments which is a fabric band worn around a the ribcage. Electrodes detect the heartbeat and the measurements are transmitted to an accompanying wristwatch and stored for later download.

Galvanic skin response (GSR), also known as electrodermal response, is a measure of the skin's electrical conductivity, which is affected by the amount of perspiration exuded. Skin conductance is measured between two electrodes placed on the skin and is measured in μSiemens. While at rest, a body's rate of perspiration is unlikely to be due to physical exertion unless we are resting in a very hot and/or humid environment. Changes in the eccrine sweat glands of the skin cause perspiration and have been linked to increased emotion, arousal or attention [2], therefore it is likely that changes and peaks in the galvanic skin response signal while watching a film are due to emotional stimuli provided we remain seated and the room temperature and humidity are at a comfortable level. Our participants wore a BodyMedia SenseWear PRO2 Armband, a lightweight and unobtrusive armband worn around the triceps of the upper right arm, next to the skin. This device combines a range of sensors including a skin conductivity sensor and each armband was configured to record at a sampling rate of 4 Hz.

Finally, the **amount of movement** of each participant during a film can indicate whether the subject may be bored, agitated, excited, shocked or surprised. Such movements are not physiological however, and can be masked to present indicators of emotional stimuli where there are actually none. For example, a participant may move in their chair in time to the beat of some film music, or may decide to change position on the seat because they are uncomfortable or feeling cramped. The chairs used by our participants had a foam-based pressure sensor integrated into its backrest to record movements and changes in the subjects' posture. The sensor is fabricated from a type of polypyrrole-coated polyurethane foam [3], the electrical conductivity of which changes when it is compressed, and this can be measured.

In total, 51 full-length feature films were shown over a ten-week period in the CDV*Plex*. The participants in our experiment were 43 volunteers from across the university's population of staff and graduate students, of which 18% were female. Film showings were advertised daily using posters and email bulletins and an online booking system and website were provided. Participants were limited to at least two and at most four film viewings to ensure that we could compare result sets for the same person yet maximise participation. The majority of viewers were in their early twenties, and in almost half the viewings, participants chose films they hadn't seen before.

In order to detect possible indicators of emotion in the heart rate, GSR and movement data for each viewing of each film by each participant, we gathered data and preprocessed it in different ways according to its differing nature. For heart rate we calculated the mean baseline and identified upper and lower thresholds of significance as fluctuations of greater than $\pm$ 10 b.p.m. For GSR data we calculated the standard deviation from the mean for each data series which revealed substantial changes, and from this we identified a threshold of significance for each individual viewing. Finally, for detecting peaks in the motion data, we first calculated the mean and normalised values to 1.0. Using a sliding window approach, for each sampled value we calculated the difference between it and its predecessor, and it and its successor and if the successive difference was significantly larger than the antecedent difference then a significant movement was deemed to have occurred.

For the purpose of this paper we concentrate our analysis of people's viewings of 6 specific films (described later) as viewed by 16 participants, and in the next section we

describe how films were analysed and tagged for emotion indicators.

3. Annotation and Emotion Tagging of Films

In order to perform a correlation analysis between our observations of emotional reactions to films and the places within those films where emotions are supposedly evoked, 6 films were manually annotated for their likely emotional content. The set of films chosen were *Casablanca, Harry Potter and the Prisoner of Azkaban, Finding Nemo, Indiana Jones and the Last Crusade, Mean Streets* and *Pirates of the Carribbean: Curse of the Black Pearl*. These were chosen as they give a good spread across several film-making eras and genres. Annotation was completed by the second author and selective parts were reviewed by the first author

We analysed the audio soundtrack of each film using several low-level audio features followed by a classification process in order to detect the presence of speech and music in the film. These features included zero-crossing rate, silence ratio, short-term energy levels and the ratio of low to high energy levels and the detection and classification process is described in [8]. The process ultimately resulted in each minute of each film being classified into five categories, namely Silence, Speech, Music, Quiet Music and Other. A summary of this speech-music classification is given in Table 1 showing the amount of time within each film in which music was detected. For example, we can see that Indiana Jones III has the highest proportion of music content at 77 minutes or 64%.

Table 1. Music-Rich Segments in Films

Film	Length hh:mm	Total Music (mins.)	Music-Rich Proportion of Film
Nemo	1:34	16	17%
Mean Streets	1:45	27	26%
Indiana Jones	2:00	77	64%
Pirates	2:15	62	46%
Harry Potter	2:14	43	32%
Casablanca	1:37	22	23%

To determine the set of tags to use to describe film emotions we looked at several starting points. In recent years, Salway has conducted work on analysing feature films using the associated Audio Description (AD) [11], a text description of on-screen content and action developed to help the visually impaired. Salway developed 21 categories of emotion based on the emotions classified by Ortony et al. [9]. Salway's method for detecting the emotional content in the AD of a film was to use indicator

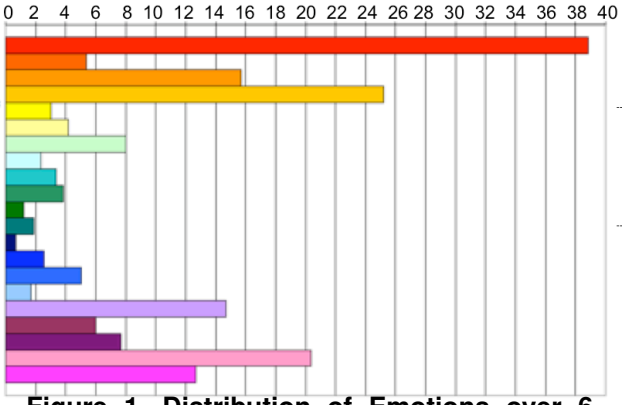

Figure 1. Distribution of Emotions over 6 Manually Annotated Films

words in the time-coded AD as entry points into WordNet [4] followed by some post-processing to ensure unique mappings from emotion tokens (words in the AD) to one of his 21emotion types or classes

We used Salway's 21 emotion types as our tags for the purpose of manually annotating the selection of films, and these are shown in Figure 1. Associated with each of these types are a number of descriptive keywords, 627 in total. Each of the six films were annotated to a per-minute granularity with one or more of 21 emotion labels and Figure 1 shows the cumulative distribution of these over the 6 annotated films. Figure 2 shows the annotation results as a time series, colour-coded by occurrence of each emotion type, for the film "Casablanca". The figure also indicates the music-rich portions of the film shaded in gray. When viewed vertically, each column indicates the combination of emotions which were present in that minute of the film, and the ordering of emotions (top to bottom) in both Figures is the following: love, joy, pride, admiration, hope, happy for, relief, satisfaction, gratitude, gratification, gloat, pity, remorse, shame, reproach, resentment, disappointment, distress, anger, hate and fear. When examining these annotation results it is worth noting

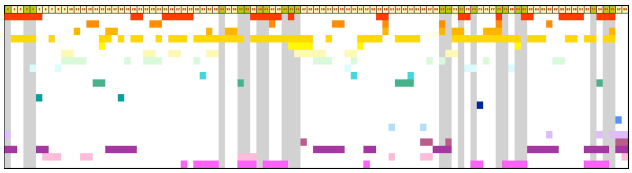

Figure 2. Distribution of emotions and Music-Rich Sections in the film "Casablanca"

that the red-orange-yellow colours represent emotions with stronger "negative" connotations such as fear, hate, anger and distress, while the pink-purple colours represent emotions with stronger "positive" connotations such as love,

joy, pride, admiration and hope. The dominance of these two sets of emotions, which can be regarded as opposite extremes, is clearly evident throughout Figure 2, where we frequently see long blocks of these emotion types.

4. Results and Analysis

The first question we investigated in our analysis of gathered data was the similarity among participants in terms of their reactions while viewing films together, as a group. For each viewing of each film, biometric data was collated, synchronised and underwent our peak detection process. From this, the proportion of potential emotionally detected peaks in each viewing was computed. The degree of similarity between these levels of biometric reaction for each subject was computed as a composite Euclidean distance measure of peak detection rates for each emotion category. In cases where there were more than two viewings of a film the similarity measure was calculated for each pair of subjects and averaged.

As an example, *Finding Nemo* was viewed by four subjects P10-f, P23-f, P28-f and P35-f, all female. The subjects' average pair-wise similarity measures are quite close, with a mean similarity measure of 91% across all pairs and Figure 3 shows the pairwise similarities and the average, across the emotion tags. Apart from HATE, DISAPPOINTMENT, RELIEF and PITY, the pairs agreed hugely on the occurrences of their biometric peaks. Figure 3 illustrates the degree of similarity between the emotional peak detections for all pairwise viewings of the film "Finding Nemo".

Table 2. Viewing similarities for *Finding Nemo*

Viewing Combination		Similarity Measure
P10-f	P35-f	95%
P23-f	P28-f	93%
P23-f	P35-f	91%
P28-f	P35-f	89%
P10-f	P23-f	88%
P10-f	P28-f	87%
Avg. Similarity		91%

Taken as a whole, these figures indicate that when people view a film together they respond in a similar manner. In all cases but one, the measures of similarity between viewers of the same film were over 90%. Table 3 gives a summary of the results of joint participant reaction to films, for each of the six films we tested.

The second topic we investigated in our analysis was whether emotional responses to stimuli in films can be detected using the sensors we used. We plotted the average number of *observed* emotion type occurrences for all

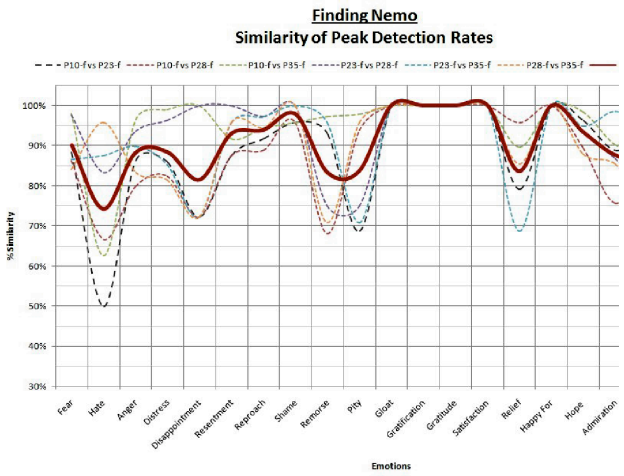

Figure 3. Similarity of Viewers' Emotional Peak Detection in *Finding Nemo*

Table 3. Summary of Peak Detection and Viewer Similarity Results

Film	No. of Viewers	Mean Similarity
Casablanca	2	94%
Harry Potter	2	96%
Indiana Jones	2	91%
Finding Nemo	4	91%
Pirates of the Carribean	3	93%
Mean Streets	3	89%

viewings against the average number of *annotated* emotion type occurrences for all viewings in Figure 4. This shows that there is a significant absolute difference between the numbers of each emotion type annotated and the absolute numbers of each emotion type detected using biometric measurements.

However, it is also apparent that there is some correlation between the two data series and when both data series are normalized into the range [0, 1] and plotted against each other as in Figure 5, the degree to which they form the same pattern of occurrence becomes immediately evident. The degree of distribution for detected peaks for each emotion type almost exactly matches that of the distribution of annotated emotion types. This result indicates that emotional responses can be detected using biometric measurements. While not all occurrences of an emotion in a film can be detected by analysing outputs of the biometric sensors, with 12% of emotion occurrences being detected on average, any occurrences that are detected will parallel the same overall emotional pattern of the film with 97% accuracy on average.

The third topic we investigated was how the presence of music affects the emotional responses detected from film viewers and in particular, which emotions are most

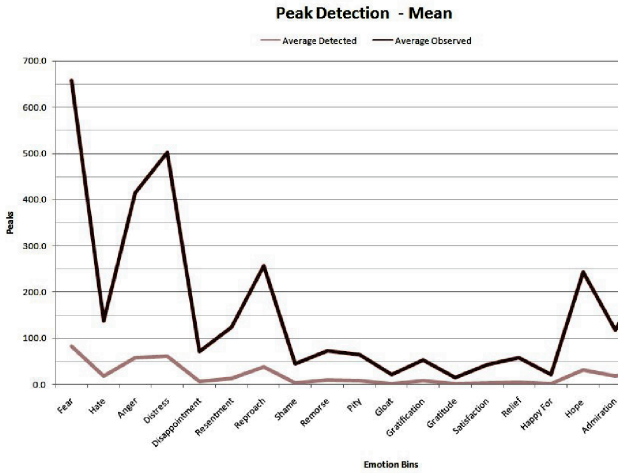

Figure 4. Overall Proportion of Emotion Occurrences Detected across all Films

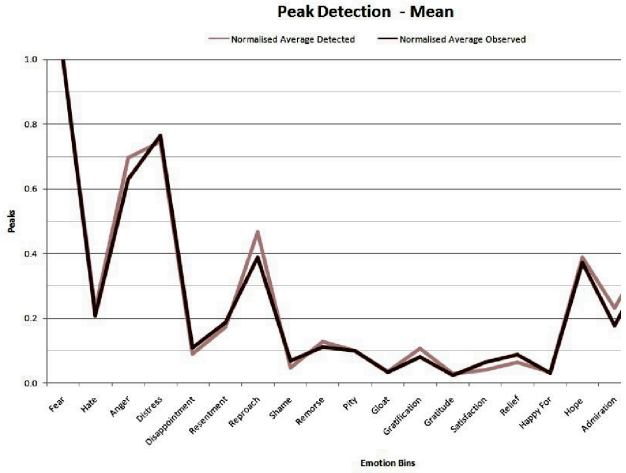

Figure 5. Emotion Detection vs. Emotion Annotation across all Film Viewings

impacted. Figure 6 illustrates the degree to which people respond to different emotions with respect to the presence of music. The data series represents the percentage effect of the average response magnitude for music from the overall average response magnitude, for each emotion type. The overall average response is therefore 0%. In this way, the degree of influence of music on the size and strength of viewers' reactions to different types of emotion is evident. The emotion types for which music has the strongest positive impact are JOY, RESENTMENT, SATISFACTION and HOPE FOR while SHAME, GLOAT AND REMORSE are the ones with strongest negative impact. This indicates that the impact of music on reaction to different emotion types is not just confined to the positive or to the negative kinds of emotion.

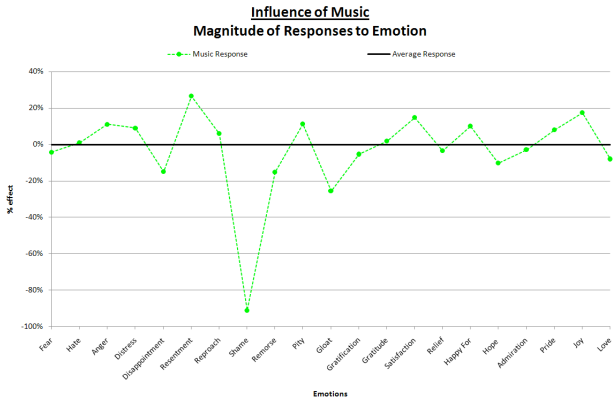

Figure 6. Influence of Music on Emotional Responses in Viewers

5. Related Work

Relatively little attention has been given in the literature to the use of physiological data in detecting emotions in people as they watch films, though some studies have been reported. Hettema *et al.*[7] describe a study of physiological reactions to film viewing but their study concentrated on differences among viewers throughout the whole of the film rather than reactions during different segments of the film to various musical and other emotional stimuli. The motivation for their study was to explore aspects of differences between personalities and unlike our work reported here there was no attention paid to music in films and the sensors used to capture physiological reactions were less complete than what we used here.

Looking beyond film and film score as emotional stimuli, Baumgartner and colleagues at the University of Zurich [1] have been studying physiological reactions to still pictures as a visual stimuli and also reactions to classical music, and especially when the two are combined. Their measurement techniques include electroencephalograms, heart rate, skin conductance, respiration and temperature. Their findings indicate that music can markedly enhance

the emotional experience evoked by affective still pictures. This finding echoes our own in that we also find similar correlations though in our case the physiological effects come from the film experience as a whole, amplified by the music and also from viewing a film in a group setting.

The emotional effects of music have also been used in a study reported in [10] where startling background music was added to TV news reports. The findings indicate that background music, even startling background music leads to higher ratings for such news stories though the effect may be either adverse or beneficial in terms of negative/positive arousal on the part of viewers. Physiological measurements of emotional states were taken using facial electromyography and electrodermal activity such as galvanic skin response, as used in our work.

6. Conclusions

Our results from an empirical approach to measuring physiological reactions to film-watching among a group of viewers, show that people experience similar and measurable physiological responses to emotional stimuli in films when they watch films in small groups. We found that the effects are more pronounced for emotions "at the extremes" such as FEAR/ANGER/DISTRESS and HOPE/PRIDE/JOY. We also found that these responses are influenced by the presence of music in the film, and that we can detect these responses through analysis of their biometric measurements. Importantly, we did not set out to see if it is the music which causes the high emotion reaction as we would have had to show films without music or otherwise manipulate the audio in order to see if it was the music causing the emotional effect. In addition we concentrated on physiological measures related to arousal, the range of responses from energized, excited and/or alert to calm, drowsy or peaceful, and not the responses that have to do with the manipulation of valence, which are responses ranging from pleasant to unpleasant. So if valence characterises the "type of an emotion and arousal is its "intensity, what this means is that we did not try to analyse the *type* of emotion but *how much* emotion people were feeling. We did this because we were interested in measuring the amplification effect of the music.

It would be interesting to expand our analysis to encompass a greater number of films across genres, perhaps discovering commonalities across films of the same genre or era in terms of emotion detection. A worthwhile further experiment would be to expand the number of viewers for each film which was just not possible in our lab setting and to test whether further biometrics, such as electromyogram (EMG), electrocardiogram (EKG), blood pressure or respiration, would improve on our emotion detection.

In terms of practical applications of our findings, the advantages are to automatic video analysis, improving video analysis and summarization techniques to enable users to search and browse films, and perhaps other fictional video, based on their emotional content. In a real world setting, this has obvious applications in film recommender systems following initial work on affective content analysis by Hanjalic [6], to allow users to find films based on the desired type of film or film experience, as well as on content.

Acknowledgment

This work is supported by Science Foundation Ireland under grants 03/IN.3/I361 and 07/CE/I1147.

References

[1] T. Baumgartner, M. Esslen, and L. Jäncke. From emotion perception to emotion experience: Emotions evoked by pictures and classical music. *International Journal of Psychophysiology*, 60(1):34–43, 2006.

[2] M. Bradley, B. Cuthbert, and P. Lang. Pictures as prepulse: attention and emotion in startle modification. *Psychophysiology*, 30(5):541–5, 1993.

[3] S. Brady, D. Diamond, and K. Lau. Inherently conducting polymer modified polyurethane smart foam for pressure sensing. *Sensors & Actuators: A. Physical*, 119(2):398–404, 2005.

[4] C. Fellbaum. *Wordnet: An Electronic Lexical Database*. MIT Press, 1998.

[5] W. Frey and M. Langseth. *Crying: The Mystery of Tears*. Winston Press, 1985.

[6] A. Hanjalic. Adaptive extraction of highlights from a sport video based on excitement modeling. *IEEE transactions on Multimedia*, 7(6):1114–1122, 2005.

[7] J. Hettema, K. Leidelmeijer, and R. Geenen. Dimensions of information processing: physiological reactions to motion pictures. *Europ. J. Personality*, 14(1):39–63, 2000.

[8] B. Lehane, N. O'Connor, and N. Murphy. Dialogue Sequence Detection in Movies. *Image And Video Retrieval: 4th International Conference, CIVR 2005, Singapore, July 20-22, 2005, Proceedings*, 2005.

[9] A. Ortony, G. Clore, and A. Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, 1988.

[10] N. Ravaja and K. Kallinen. Emotional effects of startling background music during reading news reports: The moderating influence of dispositional BIS and BAS sensitivities. *Scandinavian Journal of Psychology*, 45(3):231–238, 2004.

[11] A. Salway and M. Graham. Extracting information about emotions in films. *Proceedings of the eleventh ACM international conference on Multimedia*, pages 299–302, 2003.

[12] T. Thomas. *Music for the movies (2nd Ed.)*. Silman-James Press, 1997.