DETECTION AND ANALYSIS OF EMOTIONAL HIGHLIGHTS IN FILM

SANDRA ROTHWELL BSc.

A dissertation submitted in fulfilment of the requirements for the award of

MASTER OF SCIENCE

to the

Centre for Digital Video Processing

School of Computing

Dublin City University

SUPERVISOR: Prof. Alan F. Smeaton

September, 2007

This thesis is based on the candidate's own work, and has not previously been submitted for a degree at any academic institution

DECLARATION

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Master of Science is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

SIGNED: Sandra Pettorl

ID NO.: 99343924

DATE: 19 Sept. '07

ABSTRACT

This work explores the emotional experience of viewing a film or movie. We seek to investigate the supposition that emotional highlights in feature films can be detected through analysis of viewers' involuntary physiological reactions. We employ an empirical approach to the investigation of this hypothesis, which we will discuss in detail, culminating in a detailed analysis of the results of the experimentation carried out during the course of this work. An experiment, known as the CDV*Plex*, was conducted in order to compile a ground-truth of human subject responses to stimuli in film. This was achieved by monitoring and recording physiological reactions of people as they viewed a large selection of films in a controlled cinema-like environment using a range of biometric measurement devices, both wearable and integrated. In order to obtain a ground truth of the emotions actually present in a film, a selection of the films used in the CDV*Plex* were manually annotated for a defined set of emotions. We examine how filmmakers use devices and techniques to stimulate viewers' emotions, particularly how music is used in film to intensify the impact of onscreen action. We also examine the different event types of which film scenes are comprised and how they can be detected using audio-visual analysis of the video content. Finally, we calculate and study the correlations between the emotions found in the annotated films, and "events" or highlights in the viewers' biometric measurements, and also between the scene events and music occurring in the film and the emotions present.

To Stephen

who got me through

ACKNOWLEDGEMENTS

This work is the culmination of many people's belief in me and my ability to achieve my goals, however they may have changed over the past couple of years.

To my colleagues and friends of the CDVP; I feel privileged to have had the opportunity to work alongside such a talented group of researchers. My days in L1.31 have given me many enduring memories of friendship, comradeship…and a certain wackiness! My everlasting thanks to the gang (past and present): Sinéad, Cathal, Pete, James, Paul F., Colum, Mike, Hyowon, Neil, Fabrice, Paul B., Jiamin, Kieran, Dave, Bart and Órla. Most of all I wish to thank Georgina who, no matter what, is always there for me as my friend and counsellor. I wish to thank my supervisor Prof. Alan Smeaton for his support in difficult times. This list would not be complete without mentioning some friends who were always on hand to listen and help – for this, I'd like to thank Annette E., Dónal and Jean. To my family, Ivan, Geraldine, Lorraine and Alison, who know me better than I know myself, goes my heartfelt gratitude. They have always encouraged and supported me in whatever I chose to do. My parents' good advice and their unwavering belief in me have been my bedrock. Thank you for always listening and so much more. Last, but far from least, to Steve – saying "thank you" seems almost a paltry phrase to acknowledge his part in the completion of this work. I could not have done this without him; he picked me up when I thought I couldn't go on, his brilliant mind was my sounding board and inspiration, his constant support is the rock on which I lean. My love and gratitude always.

CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This chapter introduces the concept of detecting emotion in film, as well as providing some background information on the area of Information Retrieval and it's applications to multimedia content. We will discuss video information retrieval and the various components used within modern video retrieval systems. We will then discuss current research into improving the retrieval system's performance in returning accurate results and therefore the applications of the work described in this thesis. In this context, we will then discuss the motivation behind the concept of detecting emotions in film and, finally, we will discuss the hypotheses proposed in this thesis.

It is estimated that nearly a quarter of a million feature films, also referred to as "movies", have been made since motion pictures first appeared around 1895. Each year, thousands of feature films are produced worldwide, with over five thousand major cinematic releases in 2001 alone [Varian & Lyman, 2003]. In that year, the USA more than doubled the number released in 1991 – and almost quadrupled the number released in 1971.

Of course, of these thousands of feature films each year, we are made aware of only a tiny proportion, usually through the marketing activities of the studios. This advertising takes the form of posters and billboards dotting our urban environments, and on radios, televisions, and cinema screens appear promotional "trailers" or short clips of film footage edited in such a way as to try and hint at a film's genre and plot without giving too much away. As a result, our choice of films to watch is generally dictated by subjective marketing, or the even less objective critics' reviews. There exist movie recommender systems, such as *MovieLens* [GroupLens Research, 2006], but these rate each film based on a combination of outside metadata such as the cast, the title, and the supposed genre as well as collaborative recommendations from other users with similar interests.

While movie recommendations based on the recommendations of others are useful, common sense dictates that movie recommendations should be based partly on the actual content of the film as well. Unfortunately, the problem is that movies tend not to have much content metadata apart from the basics of title, actors, director, etc., and perhaps a brief preview trailer. These alone are not enough on which to base good recommendations and, in practice, the best recommendations come from people who have actually seen the film, but then user subjectivity comes into play once again.

Therefore the best scenario on which to base recommendations would be a completely objective summary of the film from a person who had viewed it. It is hoped that this work will provide a step towards this objective, by providing the basis for a technique by which viewers can provide a completely objective and unconscious analysis of how they are affected by the content of a film. As we will see below, this approach has the potential to enhance the field of Information Retrieval in providing a new mechanism by which video content can be indexed, segmented and summarised in a more meaningful manner for intuitive human-computer interaction in search, browsing and summarisation systems.

In this chapter, we will elucidate and define the hypotheses underpinning our work in the detection and analysis of emotional highlights in film. We look at digital video and its growth in popularity particularly in the entertainment industry, and discuss the motivation behind the use of feature films as the basis for our analysis. This chapter also introduces the area of Multimedia Information Retrieval with respect to the applications of this work in improving the state of the art in this ever-evolving and popular field of research.

1.1 MULTIMEDIA INFORMATION RETRIEVAL

Traditionally information retrieval operated over text documents from large collections, with state of the art commercial information retrieval systems or search engines, such as *Google* and *Yahoo!*, successfully searching and answering a user's information need. The development of new media technologies in imagery and audio, and integrated multiple media such as video, coupled with much improved compression algorithms and formats, have created vast multimedia libraries and archives in diverse areas such as medicine, criminal investigation, art galleries and broadcasting to name but a few. It is clear that there is a need for information management, organisation, retrieval and navigation through these vast multimedia archives. For example, the BBC archive stores an additional one million new items per year including video, image, audio and text information.

An information retrieval system which provides access to a video collection is far more complex than a traditional information retrieval system dealing with textual data alone. The main reason for this complexity lies in the inability to automatically analyse the video content accurately. The interpretation of video is more difficult due to the richness of its content including visual, audio, text and semantic information. It is the intention of this work to provide the stepping stones towards a new approach to interpreting the semantic aspect of video content though analysing how it affects its viewers, and thus identifying its emotional impact or intention.

1.1.1 Video in the Digital Age

Before briefly discussing digital video retrieval, we first need to have a basic understanding of digital video. Video is a sequence of images, displayed at the rate of twenty-five to thirty images per second, giving the illusion of motion and synchronised with an audio track. Until recently, most of this video data was in analogue form, stored on reels or tapes and transmitted through projectors, radio waves or analogue cables. Analogue video requires on average 1.3 MB of space for each image in uncompressed form. Sequential storage of this video format required a large capacity storage medium which was available only in the form of a magnetic tape in the 1970's when consumers became familiar with the concept of video with the introduction of Video Home System (VHS) by JVC. With the development of new technologies and compression standards, digital video became a reality and was introduced in the 1980's. It offered many interesting features over traditional analogue video, such as higher picture quality, easier storage and transmission across networks.

The introduction of digital media opened up a whole new range of applications for video, as well as improving on many of the previous ones. It has transformed the way in which video data is viewed, stored and transported. Video can now be stored on digital disks or hard drives, where previously analogue tapes were required. Virtually all video content now produced is available in digital format.

Digital video has also made the creation of movies and television programmes easier and cheaper as, prior to its advent, most movies and television programs were shot using expensive cameras that used reels of expensive film. Now, the cost of high quality digital cameras has reduced significantly, and unused video can simply be overwritten, rather than discarded. This can be said to have contributed to the large increase in the amount of video being created in recent years, such as the aforementioned doubling of movie output in the USA from 1991 to 2001.

In terms of research, an invaluable consequence of the use of digital, as opposed to analogue, video and audio is that it has made it possible to analyse the data. For example, digital video is represented by a set of pixel values that represent the colours which constitute the visual aspect of video. Thus, using digital signal processing (DSP) theory, it is possible to extract information from the video data and use it to gain knowledge about the content. This created the broad field of video analysis which now includes many diverse areas such as surveillance video analysis, face recognition, fingerprint recognition, object detection and object tracking.

1.1.2 Video Information Retrieval

An information retrieval system is an implementation of a software algorithm that gathers, indexes, searches and manages a document collection (text, video, image or audio) and designed with the overall aim of aiding potential users in the retrieval of information they require from the collection of data [Agosti, 2002].

Information Retrieval (IR) in the text domain has made great advances within the last decade with the improvements in capable technology and its wider availability, and the advent of the World Wide Web. Many solutions to the conventional approach of ranking relevant documents according to the degree of relevance to a query have been proposed to suit different user requirements and implemented successfully over many data sets. The main challenge within the video information retrieval research community is to achieve a similar standard of retrieval within the video domain to that which exists within the text domain. The retrieval of video data is much more complex than that of traditional text data for many reasons, including the fact that there are many more media components to be considered when manipulating the contents of a video document. Data is not only considered on a conceptual level, by issuing queries with keywords like "cat" or "computer" which is the standard form of retrieval for text documents. We are now also working on a perceptual level due to the multisensory nature of video by composing queries that also contain images, video clips or audio examples, which might contain the desired entity that the user wishes to retrieve from a video collection.

1.1.3 Challenges in Video Information Retrieval

There are some major challenges to the development of video IR including the prohibitive processing requirements associated with digital video and the difficulties in providing efficient user interfaces and video representations for browsing and search systems. Also there are the non-trivial challenges of achieving detection of perceptual events and high level concepts, such as emotional content and semantic events. Humans have an ability to easily decipher the semantic meaning of a visual image. It is far more difficult for a machine to automatically extract the semantic meaning from an image or similarly a video sequence. Within the text domain the semantic meanings are integrated into the text and text IR performs adequately at providing documents matching a user's request, however automatic semantic extraction of video remains a non-trivial and complex task. In Chapter 4 of this work we discuss a successful approach to semantic event detection in movies which was developed by our research group in Dublin City University.

In terms of the possible applications of the work proposed in this thesis, we seek to address two aspects of video IR which we will discuss below, namely indexing and video summarisation.

1.1.4 Indexing Video

Locating relevant portions of video or browsing content is difficult, time consuming and generally inefficient. Automatically indexing these videos to facilitate their presentation to the user significantly eases this process and allows for automatic information retrieval. However, creating semantic indexes which allow users to search and browse video content in a meaningful and efficient way is a task of significant complexity. Our aim is to provide the basis for an indexing methodology based on the emotions present in a piece of video, specifically movies, which would allow users to search and browse movies based on their emotional content. For example a user could search a movie for its "scary" parts or a collection of movies for all the love scenes.

1.1.5 Automatic Summarisation

Alternative retrieval applications such as summarisation systems provide users with an accurate description of the contents of the document enabling them to selectively choose documents that are most likely to answer their specific need. However, video summarisation systems are currently limited to specific narrow domains, as technology has not reached the point where all features can be accurately detected for every possible event. An example of a video summarisation system over a narrow domain is described in Sadlier, O'Connor, Murphy and Marlow [2004] where feature detectors highlight important events within soccer matches such as the scoring of a goal or a penalty being taken and it can be seen that in narrow domains, such as sport, summarisation techniques work well. A similar system which provides online access to broadcast TV news is the ANSES system [Pickering et al., 2003] developed at the Imperial College London.

Summarisation is hugely beneficial to movies, as most are at least one and a half hours long, with many as long as three hours, which means that it can be difficult to manually locate particular portions. While one approach to semantic movie summarisation is the research into automatic event detection described in Chapter 4, another approach is suggested by the work carried out in this thesis, whereby a movie could be summarised in terms of its constituent emotional content. This would allow for movies to be represented by a breakdown of the emotions it induces in its viewers. This has obvious applications for the aforementioned movie recommender systems, in that it would allow users to retrieve films based on their desired type of film or movie experience.

1.2 EMOTION DETECTION IN FILM

The work carried out and discussed in this thesis seeks to investigate the supposition that emotional highlights in feature films, or movies, can be accurately detected through analysis of viewers' involuntary physiological reactions. We employ an empirical approach to the investigation of this hypothesis, which we will discuss in detail in the following chapters, culminating in a detailed analysis of the results of the experimentation carried out during the course of this work.

1.2.1 Research Questions Addressed in this Thesis

To accurately address this hypothesis we have expounded three research questions or "mini" hypotheses

Hypothesis I

Emotional responses experienced by people watching movies can be detected through analysis of their biometric measurements for that period.

Hypothesis II

Individual viewers experience similar physiological reactions to emotional stimuli.

Hypothesis III

Movie music and filmic events in a movie do influence viewers' emotional responses.

In addressing each of these questions we will establish that emotional content in movies can be accurately detected and to what degree. We will learn in what way movie viewers are affected by the different elements that constitute a movie. This research aims to provide a basis from which improved, intuitive and more meaningful video indexing and summarisation techniques can be developed.

1.3 THESIS STRUCTURE

The remainder of this thesis is organised as follows. Chapter 2 will explore the use of music in film and the emotional impact it can create. It will discuss the psychology of music, touching on areas such as cognitive psychology, psycho-musicology and music therapy, and address how humans perceive emotion in music. The chapter will then describe how and why music is used in films, or movies, and to what affect.

Chapter 3 will describe an experiment conducted to compile a ground-truth of movie viewers' reactions to emotional stimuli in feature films, using biometric sensors in a controlled "cinema-like" laboratory. This chapter will describe the details of the biometrics used and the experimental procedure. The results of the biometric measurements will then be displayed and the process used to identify and detect physiological "events" in these readings will be described.

Chapter 4 will describe research conducted within the Centre for Digital Video Processing in Dublin City University by our colleague Dr. Bart Lehane. The chapter will describe his work in utilising automatic audio-visual analysis and feature extraction techniques in order to identify different types of semantic "events" within a movie, such as dialogue or montages, based on the film's inherent "grammar" and structure. The detection of music will also discussed in this chapter as it is one of the major components of this automatic analysis.

Chapter 5 will describe the manual annotation process used to annotate the feature films analysed in this work for the presence of a define set of emotions. The choice of emotion tag set will be discussed and the annotation procedure explained. The films chosen for annotation, and hence analysis, will themselves be described and discussed in terms of the annotation results.

Chapter 6 will present and examine the results of the experiments carried out and discussed in this thesis and describe our findings with respect to the hypotheses outlined above.

Finally Chapter 7 will conclude this thesis, where we will summarise our findings on the detection and analysis of emotional highlights in film. We will reflect on the answers to the hypothesis identified above. We will then conclude by making suggestions for further work in the area of emotion detection in biometric analysis.

CHAPTER 2

THE FILM SCORE AND CINEMATIC EMOTION

This chapter describes how music is used in feature films, and to what effect. We will explore the psychological impact music can have on the human mind. Our perception of music and its effect on our emotions is also considered, appreciation of which is fundamental to understanding how and why music is used in feature films. The art and science of film scoring will be discussed later in this chapter.

The previous chapter introduced the ever evolving field of Information Retrieval and some of its applications in multimedia. We looked in particular at work being conducted in digital video analysis and its audio-visual components. We discussed the necessity of developing systems which would allow for more intelligent browsing and search, in order to provide people with the capability of dealing with the vast and rapidly multiplying amount of video data available in the world today. To that end we introduced the concept of detecting emotional content of digital video and how this technology would aid in more logical and meaningful segmentation and summarisation methods. For the purpose of this work, feature films were selected as the video type of choice due to factors such as the quantity of reliable data available, the obvious applications in automatic recommender systems, editing and searching, and the emotion rich content. In fact, one of the most important components of feature films, with respect to this type of emotive analysis, is the almost omnipresent musical content, in our opinion.

For hundreds of years and more, attempting to explain the relationship between music and our emotional state has captured the interest of people working across a broad range of areas. However, it is the work of more recent years that we discuss in this chapter.

Advances in cognitive psychology, and the related discipline of psycho-musicology, have provided insight into how music is processed by the brain and how it can affect our neural pathways. We all realise that music does affect us emotionally, and most of us have employed music for just that effect at one stage or another. For example, people may listen to some soft music to help them unwind and relax, or play something loud and rhythmic as they go about some housework. However, psychologists seek to determine exactly how and why this phenomenon occurs.

The use of music in feature films, then, is directly related to the fact that humans are so very susceptible to the emotional influence of music. Later in this chapter we will examine how film-makers emphasise emotional highs and lows in a film's narrative, or storyline, and investigate the techniques for scoring a film. In Appendix A, we have included a summary of the key points of musical theory in order to give the reader a basic understanding of the mechanics of music.

2.1 THE PSYCHOLOGY OF MUSIC

Many have declared "music is the language of emotions" and indeed that music can profoundly affect our emotions is a well-known and universally accepted fact. However what we are less sure about is how to explain this phenomenon. For centuries philosophers, writers and musicians have argued and theorised; in more recent years however, the development of cognitive psychology has led to a more scientific interest in the field.

Cognitive psychology is the psychological science which studies cognition, the mental processes that are hypothesised to underlie behaviour. This covers a broad range of research domains, examining questions about the workings of memory, attention, perception, knowledge representation, reasoning, creativity and problem solving. It is one of the more recent additions to psychological research, having only developed as a separate area within the discipline since the late 1950s and early 1960s (though there are examples of cognitive thinking from earlier researchers). It is a way of thinking and reasoning about mental processes which considers these processes as software running on the computer that is the brain. Theories commonly refer to forms of input, representation, computation or processing, and outputs.

Psycho-musicology then seeks to fuse the findings and concepts of cognitive psychology with the intuitions that properly underpin music theory and analysis. Imagine for a moment, if we humans hadn't evolved our socio-emotional brains. Music as we know it would probably be regarded little more than interesting sequences of sounds, at worst an irritating cacophony – in any case "affectively flat". Instead, 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 eight percent of all crying episodes in the USA were evoked either directly or indirectly by music [Frey, 1985]. Less extreme physical reactions are more common however, for example "goose-bumps" and lumps in the throat. The most frequent stimuli are new or unexpected harmony, crescendos and other dynamic shifts. Females generally experience this phenomenon more often than males, with sad music being the primary source for both [Lavy, 2001]. This work also shows it is widely accepted that there is a correlation between emotional responses and physiological functioning. Common physiological responses to music are activations of the sympathetic nervous system such as changes in heart rate, blood pressure, breathing, and the aforementioned skin temperature, and the release of arousal hormones from the adrenal glands. Many of these reactions can be easily and unobtrusively monitored, as will be discussed in later in this work. Rickard proposes an "arousal hypothesis", whereby the intensity of emotion is determined by the level of this physiological activation [Rickard, 2004]. In his study, Rickard also found that these responses were significantly stronger for emotionally powerful music in comparison to emotionally powerful video, even when both were subjectively chosen.

The success of Music Therapy in helping people with psycho-physiological problems ranging from autism to Alzheimer's provides further evidence of how music affects the psychological and physiological condition. For example, AIT (auditory integration training) is a popular music-based treatment for early childhood autism which is proving to be very effective, one reason being that music tends to promote certain brain neurochemical activities, which facilitate attentional processes [Rimland & Edelson, 1995]. It is reported to reduce auditory sensitivities and produce a wide range of improvements in behaviour, from increased attention to sociability. Of course people have been making use of the healing and therapeutic effects of music since time immemorial. In ancient Greece it was considered a conventional psychiatric treatment and even today it is still recognised as being capable of easing pain, relieving stress and alleviating feelings of depression.

2.1.1 Perception of Emotion in Music

It is hard to believe that any medium can so readily evoke basic emotions of the brain, and more. We can rapidly convey levels of love, devotion and empathy through music that would be hard to achieve with any other mode of communication except for perhaps physical touch itself. And so psychologists question, what is it about the human mind that make it such a receptive vessel for the emotional power of music?

It is still not known to what degree emotional responses to music are due to properties such as rhythm and pitch, in comparison to personal memories and learned cultural significance. However, psychologists agree that these factors must interact in a complex and subjective manner in order to influence music's emotional impact. Some believe that as well as utilising cultural codes of emotional communication, music has a deeper trans-cultural, emotional aspect; that humans are so sensitive to emotional sounds because we were evolutionarily designed to recognise the survival benefits of subtle emotional communications [Panksepp & Bernatzky, 2002]. On such "fundamental emotional capacities" complex musical cultures can be constructed, which go far beyond simple evolutionary explanations.

There are distinct brain systems that control anger, fear, joy and sadness, as well as other social emotions. The more "emotionally resonant" (Dionysian) right hemisphere of the brain has been found to have a critical role in affective musical appreciation and expression, suggesting that there is an intimate relationship between affective and musical processes in the brain. The more "analytical-intellectual" (Apollonian) left side of the brain seems to control the skills needed to process music in more cognitive manner, but these skills are not essential for simply enjoying music. When the composer Ravel suffered a left hemisphere stroke, his appreciation of music remained undiminished but he lost the ability to render it into any standard notational system or performance [Cytowic, 1976].

Further, the fact that music can evoke physiological reactions, as previously discussed, indicates arousal of primitive subcortical regions of the brain which are essential for generating affective processes. In fact, it has been found that "music probably has more direct and powerful influences on subcortical emotional systems than do visual arts" [Panksepp & Bernatzky, 2002]. The sound of a high-pitched sustained crescendo, such as a note held by a soprano's voice or on a violin, seems to be the ideal stimulus, but any solo instrument emerging suddenly from a soft orchestral background is also very effective. There is a theory that this phenomenon may trigger socio-emotional systems that generate separation-distress, based on the fact that sad, chill-inducing music contains acoustic properties similar to the separation call of young animals common to all species – the sound of someone in distress, especially a child, may make us feel cold, sending shivers down our spine.

Cognitive psychology has tended to focus on music of diatonic melody and harmony, and associated rhythmic elements, and therefore little, if any, research has been conducted using more "unusual" music of Renaissance or Medieval polyphony, Baroque counterpoint, late Romantic chromatic harmony or mid-twentieth century compositions of astructural "Serial" music and random structuring. Similarly, there is little work using Afro-American music, despite its rise to prominence since the 1950s. The implication is that diatonic structures of the late eighteenth- and early nineteenth-centuries somehow ideally matched brain structures, implying that other musical expressions might be "aberrations" [Walker, 2004]. History demonstrates that the effects of Western tonal music on behaviour were invented and deliberately practiced by musicians. Therefore, human behaviour is affected by Western music because we have assimilated the beliefs and intentions of the western cultural traditions in music.

In the eighteenth-century, Friedrich Marpurg compiled a list of the "Acoustical Expression of Emotional States", a categorisation of mood states and emotions and the types of music that can evoke them. Rather antiquated in style, it is still recognisably relevant today; though as can be seen in Table 2.1(a), most of these descriptions are still very inexact and subjective.

In some cases Mozart's accessible melody and harmony have been found to stimulate learning, clarity and creativity; Haydn's "Apollonian" orderliness can dispel irritation and depression whereas Wagner's extravagant "Dionysian" style can inflame intense emotional experiences that will either profoundly move or overwhelm and repel. Mozart has also been found to increase alertness and mental organisation, while jazz or Romantic music can have the opposite effect by loosening up mental processes.

Emotion	Expression
Sorrow	Slow, languid melody; Sighing; Prevailing dissonant harmony
Repentance	The elements of sorrow, except that a turbulent, lamenting melody is used
Happiness	Fast movement; Animated and triumphant melody; Warm tone colour; More constant harmony
Contentment	A more steady and tranquil melody than with happiness
Hopefulness	A proud and exultant melody
Laughter	Drawn out, languid tones
Fear	Tumbling downward progressions, mainly in the lower register
Fickleness	Alternating expressions of fear and hope
Timidity	Similar to fear, but often intensified by an expression of impatience
Love	Constant harmony; Soft, flattering melody in broad movement
Hate	Rough harmony and melody
Envy	Growling and annoying tones
Compassion	Soft, smooth, lamenting melody; Slow movement; Repeated figures in the bass
Jealousy	Introduced by a soft, wavering tone; Then an intense, scolding tone
Wrath	Expression of hate combined with running notes; Frequent sudden changes in the bass; Sharp violent movements; Shrieking dissonances
Modesty	Wavering, hesitating melody; Short, quick stops
Daring	Defiant, rushing melody
Innocence	A pastoral style
Impatience	Rapidly changing, annoying modulations

Table 2.1(a): Marpurg's Acoustical Expression of Emotional States

Genre	Impact
Slower Baroque (Bach, Vivaldi)	Security, precision, orderliness
Classical (Mozart, Haydn)	Lightness, visionary, regal, perception
Romantic (Tchaikovsky, Chopin)	Emotion, warmth, pride, romance, patriotism
Impressionist (Debussy, Ravel)	Feelings, dream images, day dreaming
Afro-American: Jazz, Blues, Reggae	Joyous, heartfelt feelings, sly playful, ironic
Latin: Salsa, Rhumba, Samba	Sexy, heart-pounding, body stimulating
Pop, Big Band, Country & Western	Centred, feeling of goodness, contained movement
Rock	Aggressive movement, building or releasing tension
Heavy Metal, Punk, Grunge, Rap, Hip Hop	Animating the nervous system, rebellious behaviour

Table 2.1(b): Sonnenschein's Effects of Some Musical Genres

Fast, loud rock 'n' roll can cause shallow, fast breathing, superficial scattered thinking leading to mistakes and accidents, whereas the opposite effect is produced by calming Gregorian chant. Sonnenschein [2001] suggests a list of musical genres and some of their physical, mental and emotional effects for which they are used by music therapists, a section of which is shown in Table 2.1(b). Though by no means exhaustive, it gives an indication of the impact different types of music can have on listeners.

In order to identify the specific properties of music which may be responsible for such effects, music psychologists interested in emotion have conducted a plethora of experiments. These have successfully demonstrated that for those individuals familiar with Western idioms, the perception of the emotional expression of music is based on parameters involving mode (key), tempo, intensity and timbre.

Mode

The most influential of these musical parameters seems to be mode, as diatonically major keys are consistently associated with positive emotions (upbeat and happy), and minor keys with negative ones (sad and plaintive). Indeed the ancient Greeks felt that playing music in a particular mode would incline one towards specific behaviour associated with that mode.

Tempo & Intensity

Tempo also has a significant effect on emotion, with fast music associated with joy and activity, and slow music with sadness or solemnity. Of course the effect of tempo is intertwined with and influenced by that of other parameters such as the aforementioned mode, melodic contour and particularly intensity which while not independently having any specifically defined emotional association, heavily influences the perception of tempo. The intensity of a sound equates to its energy, that is, how loud or soft it is. A high intensity will make fast music appear more energetic than its low intensity equivalent; it will make slow music more serious or solemn, whereas a low intensity makes it appear sadder, more introspective and gloomy.

Timbre

As previously discussed, an instrument's "timbre" may be said to be the result of the filtering effect of the resonating body of the instrument. The effect of the tone or "colour" of the music is complicated and far reaching, but it has been consistently found that the tonal quality of music affects perception of emotional tension.

Momentary Responses

It is worth noting that the above are temporally long-term parameters of music, which tend to remain constant throughout a piece, i.e. a musical work is slow or in a minor key etc. However, emotional response to music can also be intense and momentary, as the listener reacts to the dynamic qualities of music as it changes over time, melodically and harmonically building and resolving tension [Lavy, 2001]. It has been found that listeners generally detect melodic and harmonic structures and timing deviations in music, and their associated reactions correlate with the perception of the piece's emotionality. The dynamics of tension building and resolution generate in the listener a constant flux of expectancies which are either fulfilled or violated as the piece unfolds. Expectancy violations are particularly prone to evoking emotional reactions because they are aroused when a tendency to respond is arrested or inhibited, and in music, unlike in the real-world, this inhibition is significant because the relationship between the tendency and its resolution is made explicit and apparent.

2.2 MUSIC AND THE MOVIES

> *"Music creates order out of chaos, for rhythm imposes unanimity upon the divergent, melody imposes continuity upon the disjointed, and harmony imposes compatibility upon the incongruous"*
>
> Yehudi Menuhin, master violinist of the 20th century

In 1944, when composer David Raskin enquired as to why director Alfred Hitchcock wasn't planning to use any music in his latest production, *Lifeboat*, he was told that the director felt that since the entire film is set in a lifeboat on the open ocean, "where would music come from?". The composer quickly replied, "Ask Mr. Hitchcock to explain where the cameras come from, and I'll tell him where the music comes from."

Despite such disagreements, music and film have continued to be partnered today, as they have from the very earliest days of moving pictures. There is a historical precedent for music and drama working together that reaches back into the 19th century and beyond. The sometime lack of understanding displayed towards its use, as demonstrated by even the great Alfred Hitchcock, is perhaps because music is the most abstract of the film arts, subconsciously stimulating and guiding emotional responses to the visuals.

Music always affects how the audience will perceive a scene – we see an actor shout, but it is the music which tells us that he is joyful, not angry. Like lighting, free of verbal explicitness, music sets moods and tonalities in a film, but it differs from lighting and other filmic elements in that we hear it as opposed to seeing it, and hearing is less direct than visual perception. Therefore most films relegate music to the viewer's sensory background, as we tend to focus attention on the narrative and visual realities on the screen. However, music's influence can be felt on many levels, from its purely functional roles in serving as a "cohesive force" [Thomas, 1997] and storytelling device, to its more elusive role in creating the atmosphere and colouring the tone of a film. As composer Aaron Copland once said, "I wish more audiences could have the experience of watching a movie without any music and then seeing it the second time with music added. I think that would give them a full sense of what music does for making the cold movie screen seem more humane, more touching, more civilised".

Music draws us into the world of make-believe, the diegesis the filmmakers propose, immediately forging a connection with us on any and all of its many levels, in a way that the detachment of a silent progression of images can not emulate – the viewer is no longer simply a bystander but is coaxed into the illusion, or in some cases grabbed and hurled head-over-heels.

2.2.1 Historical Perspective

At the turn of the 20th century, in the very earliest days of filmmaking, silent "movies" wowed their first audiences, and yet it is often overlooked that they were rarely shown without some form of musical accompaniment, ranging from a single piano to a full orchestra. Even before 1910, some films had specially composed scores written for them.

Using music to tell a story was not a new idea in the entertainment business, and the use of music in films evolved out of the prevalent use of elaborate music in nineteenth-century popular theatre, to the point where "words had lost their necessity" [Gorbman, 1987]. In fact, the immediate forerunners of film music can be found in the melodramas of the nineteenth-century, to which many of the clichés of film music can be directly traced.

In the "new" medium of cinema, musical accompaniment provided an emotional context in which to place the much touted moving images. This music aided the narrative in depicting the historical, geographical and atmospheric setting, identifying characters, qualifying actions and setting the pace.

In the times where plot exposition was restricted to intertitles (on-screen text), music was the glue that held the story together. It continues to be so today, though not quite as overtly, in that it helps to maintain a sense of continuity over spatially discontinuous shots or to bridge between disparate scenes. With the advent of sound technology in the 1920s, and the miracle of speech, the role of music was shifted into "a much more intimate and much more emotionally effective part" [Gorbman, 1987]. Despite huge advances in editing, camera work & cinematography, music is used in the sound film today as it was then, to give depth, rhythm & "life" to the pictures.

2.2.2 The Effect of Film Music

Most of the time the movie-goer is not explicitly aware of a film's incidental music, yet it communicates subliminally with our subconscious, only occasionally making itself known to us as it overwhelms the senses with emotional significance, only to subside once more. However, it works away in the viewer's sensory background which, according to psychoanalytic theory, is "least susceptible to rigorous judgment and most susceptible to affective manipulation" [Gorbman, 1987]. To put it another way, music "does not pass through the same control circuits, because it is almost directly plugged into the psyche".

Likened to hypnosis for the way in which both induce a kind of trance, filmmakers have long exploited the harmonic, rhythmic, melodic suggestiveness and channelling effects of music for cinema, the object being to lull the audience into being less critical of the fantasy that is the diegetic or on-screen world. When music is working in this way, the audience is usually not consciously aware of what's going on musically in the background. It can create a dynamic subjective bond between the action onscreen and the viewer, which moves between involving them in the narrative and the emotions of the characters, and externalising them from the narrative and unifying the emotions of the audience. The former is known as "identification" music, and strives to draw the viewer into the action, to be subtle in its execution and to be intimate. On the other hand, the latter "exhibition" music is there to be noticed. It is used to punctuate the flow of narrative, to externalise the viewer and place them in contemplation of the story as a spectacle. It is used to impart an "epic" quality on the action it accompanies.

What, Where & Why

Functioning in a purely musical sense, film music can create tension and resolution through its own intrinsic and involved structure and syntax, though this is always subordinated to the narrative's demands. The score is typically composed in short phrases, which gives a maximum flexibility of musical form.

All music carries cultural associations, most of which the music industry have further pigeonholed and codified. We all know romantic music or medieval music or Native American ("Indian") music when we hear it. Instrumentation, rhythm, melody and harmony combine to create a signature which should be instantly recognisable to anyone raised in the idiom of Western music, some arbitrary examples are given in Table 2.2(a) below. This is also why film music has remained rooted in the late-Romantic style of composers such as Wagner and Strauss, as its main goal is to impart meaning to a broad audience quickly and efficiently.

What music conveys to the audience is also influenced by the filmic context in which it is used, that is, its cinematic relationship with elements in the film, major examples being beginning and end titles, and musical themes.

So while the film score, as music, has its own signification, it also has a relationship to the film in which it appears; that is, it can either parallel or resemble the action or mood onscreen, or counterpoint or contradict it. The choice of music for a film sequence will have an effect on how the audience perceives the scene. As can be easily demonstrated by simply taking a section of a film and applying different types of music to it, different combinations can have vastly different effects depending on the dynamics and structure of the music.

Place/Time/Characterisation	Musical Signatures
Middle-East	Woodwind instruments, minor-key melody with much ornamentation
Native-American ("Indians")	A 4/4 allegretto drumbeat (or bass pizzicato) with the first beat strongly accented, with a simple minor modal tune played by high woodwinds or strings
Latin-America	A rhumba rhythm and major melody, trumpets and instruments in the marimba family
Japan or China	Xylophones and woodblocks playing simple minor melodies in 4/4 metre
Turn-of-the-century Europe	Strauss-like waltzes
Rome and Paris	Accordions
Middle Ages or Renaissance	Harp music
City setting (e.g. New York)	Rhythmic, jazzy or slightly discordant major theme on brass instruments or strings.

Table 2.2(a): Examples of Geo-Cultural Associations of Music

A theme is a musical element which is heard more than once over the course of a film and as such becomes associated with narrative elements such as a character, place, situation or emotion; it can be fixed and static, or it can develop with the plot to reveal a deeper layer of meaning. It includes tunes or melodic fragments, distinctive harmonic progressions, even "theme songs". Generally, a thematic motif is not only used as a denotative label, but can also suggest meaning through its cultural connotations. The music-referent relationship can be established though dialogue, or through camera-action such as the close-up, which singles out the subject with which the music will be associated.

The effect of the absence of music from a scene is often underestimated; it can sometimes be the most powerful sound of all. Diegetic sound with no music can make what's happening onscreen seem more immediate, more tangible, causing the viewer to focus more on the images. It can also help to emphasise the fact that the characters are not talking. Further, convention has made an anchor of non-diegetic music, so that its absence from an emotionally ambiguous scene can present the audience with an image they may not interpret completely. Complete non-diegetic silence is sometimes used to depict intense mental activity, such as dream sequences. It can also have comedic uses. In any case, the transition from sound to silence, and vice versa, is itself a major factor in how we perceive that silence.

When music remains absent from a film for more than a period of about fifteen seconds, the link to the previously established tonal centre is lost. That is, within this time limit our brains will retain the musical reference of the tone and chords. So if the composer wants to change the music, fifteen seconds of musical silence is what it takes for our brains to "forget" what came before and be ready to accept the change in musical direction. Breaking this constraint creates a jarring effect, which may or may not be the desired reaction. A unifying effect can be achieved by not having to utilise this break-up effect by having the music modulate and evolve rather than change abruptly.

The use of songs with lyrics in a movie, in some respects, requires the narrative to yield to spectacle. Rather than participate in the action they can provide a chorus-like commentary on the story.

Storytelling

What are the elements of a good story? Plot and character obviously, and a strong sense of beginning, middle and end, but of course these mean nothing if there is no emotional bond with the audience. When silent films first began to give way to sound, dialogue at first retained the emotional excess that was the norm for intertitles, but it was found that audiences' reaction to the "mushiness" of the words when spoken rather than written was quite different. A reviewer in an issue of *Variety* from 1929 observed that "someone in the audience titters and it's all off. Hereafter the love passages will be suggested with the romantic note conveyed by properly pitched music" and the narrative ability of music has been exploited by Hollywood ever since. Music is commonly used to give emotional depth to characters, and give indications of personality type.

As previously discussed, the type of music used in most Hollywood productions has very strong connotative values. By interpreting the image, music can provide narrative cues, thereby giving advance notice of a threat (as demonstrated very effectively in *Jaws*) or a setup for a joke, and it can help "anchor" the audience in the story, using these "codes" to illustrate settings, characters and narrative events. "It interprets the image, pinpoints and channels the 'correct' meaning of the narrative events depicted" [Gorbman, 1987]. It can suggest changes in point of view and it can ensure continuous, often narratively "illogical" progression from one viewpoint to another.

Gorbman also notes that "music functions as connecting tissue – a provider of relations" in that it helps to smooth over any gaps, cuts, camera movements and distracting diegetic silences, and lessens awareness of any discontinuities. This is often most evident in montage sequences, where music is used to bridge gaps of diegetic time, space and situation. Music can play a vital role in maintaining many different types of continuity, including temporal and spatial, as previously mentioned, as well as thematic, dramatic, rhythmic and structural.

Film music has been described as the "emotional weather" of a film in that it is "programmed to match the mood or feelings of the narrative scene of which it is a part" [Gorbman, 1987]. The use of music in a film sequence can add "empathetic value" [Sonnenschein, 2001], emphasising the emotion the filmmaker wanted to communicate to the audience. Sometimes, in order to create a sense of irony or to establish a certain detachment from the events onscreen, the music used will not relate to the emotion of the action and as such is considered anempathetic.

2.3 THE CLASSICAL HOLLYWOOD MODEL

There exists a "pool of conventions" for classical film narrative which can be combined in different ways; a model determining a film's duration, possibilities of narrative structure, cinematography, editing and sound recording and mixing. Even allowing for the diversity of genres, directorial and authorial styles, there is something identifiable as classical Hollywood cinema. Classical cinema is said to be predicated on "telling a story with the greatest possible transparency" [Gorbman, 1987].

This classical "mode" is typified by visual editing conventions motivated by dramatic logic, accommodating the viewer's need to see details of narrative importance; spatial intelligibility is ensured by policies such as the 180-degree rule, the eyeline match and the shot reverse-shot pattern [Bordwell et al, 1985]. While the classic format was at its most specific in Hollywood feature films of the 1930s and 1940s, its features have dominated commercial cinema to the present day.

Use of background music is a major component of classical cinema, and is found across almost all movie genres from melodrama to science-fiction, war and adventure to comedy. Principles similar to those motivating classical editing practices "underlie the composition, mixing and audio-visual editing of film music" [Gorbman, 1987].

2.3.1 The Five Principles of Classical Film Music

Inaudibility

Music is not meant to be heard consciously and should subordinate itself to dialogue and visuals. There are several practices motivated by this principle. The duration of a music cue is determined by the duration of the visual sequence. Thus film composers are encouraged to build pauses and sustained notes into the music to accommodate changes made to scenes in editing. For the same reason, composing in short musical phrases and use of sequential progressions (where each restatement of a motif begins a tone or third higher than the last) are also advised. Sequential progressions can be used effectively to build tension.

Music is subordinated to the voice by lowering the volume of music behind dialogue. Also common is the practice of keeping the pitch of music away from the pitch-range of the voice, for example bass instruments used against high voices and vice versa.

Points at which music enters or leaves the soundtrack are also important. Typically, music enters or exits on actions (actor's movement, a closing door) or on sound events (a doorbell, a phone ringing), or at a decisive rhythmic or emotional change, or under dialogue - though it almost never enters simultaneously with the entrance of voice as it would drown out the words.

Referential Narrative Cueing

Music gives referential and narrative cues, indicating the point of view, referencing formal boundaries and levels of narration, and establishing setting and characters. For example, music usually accompanies the opening and end titles of a film. Music played over the opening title defines the film's genre and sets a general mood. It often states one or more of the themes to be heard later accompanying the story, setting up expectations of the narrative events to follow. Ending music tends to appear in the final scene and continues or modulates behind the end credits. It often consists of a rising orchestral crescendo with tonal resolution, sometimes involving a final statement of the score's main theme.

Using well-established conventions of harmonies, melodic patterns, rhythms and orchestration, music contributes to the story's geographical and temporal setting. Music may be associated with a particular character's point-of-view in order to emphasise it.

Connotative Narrative Cueing

Music expresses moods and connotations which aid in interpreting narrative events and indicating moral/class/ethic values of characters. Attributes of melody, instrumentation and rhythm imitate or illustrate onscreen events. Music reinforces what is usually already signified by dialogue, gestures, lighting, colour, tempo of activity and editing.

"Mickey-mousing" is the practice of imitating the direction and rhythm of onscreen actions, for example near the beginning of *Casablanca* a fighter plane is shot down and the score imitates its fall to the ground. A "stinger" is a musical *sforzando* (sudden loudness) used to illustrate sudden dramatic tension. Silence can also "sting", for example a quick crescendo to a high, dissonant chord cutting abruptly to silence.

Emotional Signification

Music may set specific moods and emphasise particular emotions suggested in the narrative (as above), but beyond such emotional connotations music is also a signifier of emotion itself.

Continuity

Music provides formal and rhythmic continuity between shots, in transitions and by filling "gaps". It can be used to smooth the visual, spatial and temporal discontinuities within scenes and sequences. Montage sequences are almost invariably accompanied by music. It is also used to bridge the gap between scenes or segments, and for spatiotemporal transitions. For example, consider a shot of a man leaving a house, a shot of the front of the library, a shot of the man flicking through a book, then a shot showing him sitting at a row of desks – all accompanied by one uncut piece of music.

Unity

By repetition and variation of musical material (tonal relationships, themes) and instrumentation, music aids in the construction of formal and narrative unity.

2.4 STUDYING THE EFFECT OF MOVIE MUSIC

This chapter describes how and why film music is used in the movies. The ultimate aim of the work carried out in this thesis is to detect and analyse emotions in movies, and in Chapter 1 we introduced three hypotheses which we want to investigate. The third of these research questions relates to film music and whether it influences how people react to emotions while viewing a film. The research conducted in this chapter strongly suggests that music has a significant effect on how people perceive emotion in general, and that film composers and music editors exploit this in how they utilise music in feature films. We will further investigate this assertion in the coming chapters.

In order to study the effects of music in an experimental sense we must first be able to detect when and where it is used. Chapter 4 describes research recently conducted in audio-visual analysis of feature films which, in part, involved the automatic detection and classification of music present on the soundtrack of a digitalised movie. This work will allow us to identify music-rich segments of a movie for further analysis. However, in order to investigate how music influences viewers' perception of emotion, we must first have a means by which to identify how viewers respond to emotion in film and whether their reactions can be captured and detected in an accurate and measureable manner. This challenge will be addressed in the coming Chapter 3.

CHAPTER 3

THE CDV*PLEX* BIOMETRIC CINEMA

This chapter describes the CDVPlex, an experiment conducted in order to compile a ground-truth of human subject responses to stimuli in feature films. This was achieved by monitoring and recording the physiological reactions of people as they viewed a feature film in a controlled cinema-like environment. This chapter introduces the set of biometric features we monitored and then describes the cinematic environment, known as the CDVPlex, which we constructed in order to allow subjects to view movies while having their biometric responses monitored. The results of the biometric analysis and the peak detection processes are then presented and explicated.

The rationale behind the use of feature films as an emotional stimulus has been considered in some detail in the previous chapter. In essence, the motivation is that conventional feature films, by their very nature, seek to capture and hold their audiences' attention by manipulating their emotional state to mirror the highs and lows of the narrative, or story, through devices in their rich audio-visual content. Not least of these is the accompanying score or musical soundtrack, which has an important and largely sub-conscious role in influencing the level of emotional commitment, and hence reaction, that a viewer experiences with the film.

Research in the field of psycho-physiology [Cacioppo et al, 2000] shows that there is a strong relationship between physiological reactions and emotional states. According to researchers in the areas of Health and of Sports Science, human physiological responses to emotional stimuli and to physical exercise are quite similar [O'Gorman, 2005]. What this means for us is that because there is a well-developed set of techniques for measuring biometric well-being during exercise, we are able to utilise and adapt a selection of these established biometric measurement techniques that were originally developed for that purpose. These biometrics will be discussed in the next section along with descriptions of the sensor technology used to detect and record the human physiological signals. We will then describe the experimental set-up and procedure for measuring biometric responses while viewing movies, including details of the volunteer participants who acted as subjects for biometric monitoring and the viewing matter, namely the set of movies shown in the CDV*Plex*. Finally, details of the biometric data peak detection will be specified and discussed.

3.1 SENSING PHYSIOLOGICAL REACTIONS

Within the disciplines of Health and of Sports Science it is known that emotional responses to stimuli and physical responses to exercise are, physiologically, quite similar. This probably comes about because exercise releases endorphins into the blood stream which creates a sense of elation and emotional high, though this is often masked by the physical exhaustion we sometimes feel when we finish taking exercise. In general though, taking exercise gives us an emotional high as well as raising our heart rate and increasing our blood circulation. In the context of the work described in this thesis, we are able to exploit this as parallels can be drawn between certain biometric measurements of both taking exercise and of responding emotionally to stimuli, and techniques developed and used extensively in the area of exercise-based research can be adapted to detect periods of emotional stimulation in people using the same biometric signals.

Relatively little attention has been given in the literature to the use of physiological data in detecting emotions in people, compared to the use of audio-visual methods such as facial expressions and speech. This is despite the fact that there are advantages of being able to gather continuous data from physiological measurements from a participant that is largely independent of the possible "social masking" that humans can employ, such as consciously controlling our facial and our vocal reactions to stimuli. Physiological reactions cannot be masked in the same way speech and facial expressions can as they are controlled by an autonomous nervous system. One of the main reasons for the under-use of physiological signals in emotion recognition is that it is difficult to map physiological reactions onto specific emotion types. It is intended that with the research presented in this work, a step towards a possible solution to this problem will be made evident and indeed this is the motivation underpinning our first hypothesis as presented in Chapter 1, which states that emotional responses experienced by people watching movies can be detected through analysis of their biometric measurements for that period.

For the purposes of the experiments we carried out in this thesis, the medium we used to induce this emotional stimulation in volunteer participants is video, specifically, a range of feature films or movies of several different genres. These will be discussed further in the next section.

3.1.1 Biometric Measurements

We chose a number of biometric measurements for our experiments, which are detailed below. The choice of measurements were based both on the recommendations of researchers in the Centre for Sports Science and Health in Dublin City University [O'Gorman, 2005] and the need for the sensor devices to be unobtrusive and if possible, wearable. This was to ensure that the data for physiological measurements was obtained under as natural as conditions as possible, while still being in a controlled, though admittedly unusual, laboratory setting, the details of which will be given in the next section.

Heart-Rate

Heart-rate is a measure of how many beats per minute a person's heart is producing at any given time. A person's heart-rate will increase as a mechanism for the body to pump more blood through the lungs and around the body. This usually happens when we need more blood to nourish our muscles as we are undergoing some form of increase in body activity, or as a response to hormones because we are stimulated, afraid, experiencing a sudden shock, or other emotional stimulus. Measurement of heart rate is a very reactive metric and responds quickly to stimuli. To make a generalised assumption, fluctuations of greater than $\pm$ 10 b.p.m. from an individual's baseline value can be interpreted as such a response. Similarly, anomalies such as peaks and plateaux in the signal waveform can indicate that the subject was experiencing an emotion or change in emotional state. Also, large spikes in the heart-rate graph such as jump of approximately 50% over the baseline, returning to normal within one minute, can be 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 approximately 3 to 4 minutes can be an indication of a period of continuous excitement.

Galvanic Skin Response

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 we exude. 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 and breathing alone causes perspiration). Changes in the eccrine sweat glands of the skin are known to cause perspiration and have been linked to measures of emotion, arousal and attention [Bradley, Cuthbert & Lang, 1993], therefore it is likely that changes and peaks in the galvanic skin response signal during this experiment are due to emotional stimuli since our volunteer subjects remain seated while watching the movie. While not as reactive as heart-rate, the galvanic skin response metric can nevertheless indicate periods of emotional change.

Motion

Motion is a measure of how often the participant moves, and the manner in which he/she does so. This information can indicate outward physical behaviour of the subject which can be useful in a number of ways. For example, the amount of movement of each participant in our experiments can help to indicate whether the subject may be bored (not much movement – slumped back in the chair), agitated (restless, frequent movements – shifting about in the chair), excited (sitting or leaning forward – on the edge of their seat), shocked or surprised (sudden, isolated movement – subject jumps in their seat). These movement measurements are not physiological however, and like other non-physiological metrics such as facial expression and speech can be masked to present indicators of emotional stimuli where there are actually none. For example, a participant may move in the chair in time to the beat of some movie music, or may decide to change position on the seat because they are uncomfortable or feeling cramped, or most likely, a participant may move because another participant may say something to him/her, to comment on a section of the movie ("wow, did you see that?"), and our participant may have to move to lean over or turn in the chair in order to hear.

3.1.2 Sensor Devices Used

A range of bio-sensors were employed to measure and record the biometric measurements introduced above for each of the participants as they watched one of the movies. These sensors were integrated into devices which utilised technologies both old and new, established and experimental. Two of the devices were portable and wearable while the other was embedded into the cinema furniture in the modified CDV*Plex* laboratory.

Polar S610i™ Heart-Rate Monitor

The heart rate monitor is a popular, conventional device used by athletes and researchers alike and consists of a fabric band which fits around a person's ribcage with a transmitter positioned in the middle of the chest area. The electrode panels in the band detect the heartbeat and the measurements are stored by transmitting the data to the accompanying wrist-watch, shown in below in Figure 3.1(a). The sensor necessitates the chest-band being worn next to the skin, beneath clothing, which makes this sensor both wearable and unobtrusive.

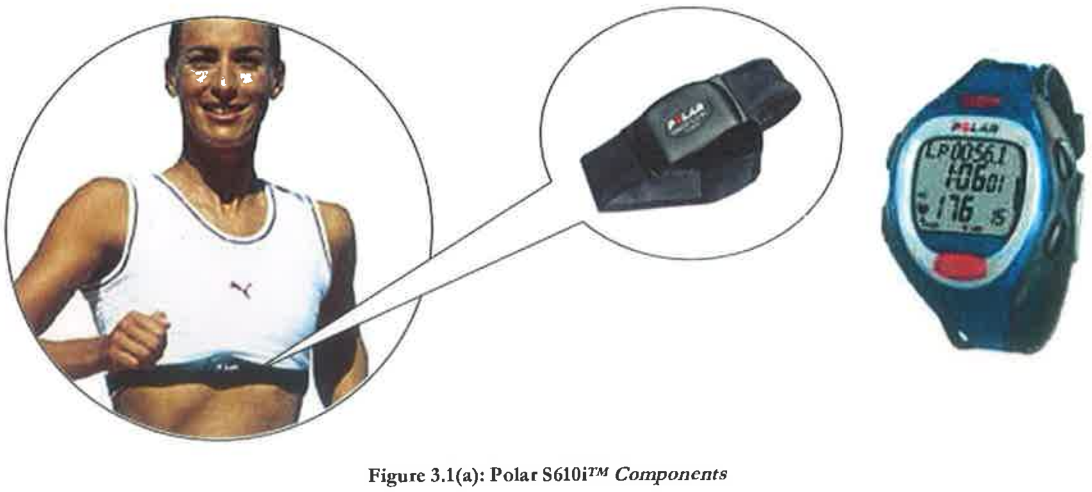


Figure 3.1(a): Polar S610i™ *Components*

The data can be time-stamped by pressing the red button on the wrist-watch, which adds a tag to the log file which is useful in marking significant events such as the start and end times of the feature film.

BodyMedia SenseWear® PRO2 Armband

The BodyMedia sensor array is in the form of a light-weight and unobtrusive armband worn around the triceps of the upper right arm, next to the skin. The BodyMedia device encapsulates a range of sensors, including the skin conductivity sensor which is relevant to this experiment. Each armband was configured to record and store up to 6.5 hours of GSR data as detected by the onboard sensor at a sampling rate of 4 per second. A tag could be added to the log file as a time-stamp recorded by pressing a button on the outer casing of the armband which, along with the location of the rest of the sensors, is shown the diagram below, Figure 3.1(b).

Figure 3.1(b): BodyMedia SenseWear® PRO2 Armband

Smart Chair Technology

Each of the chairs used by our volunteer participants while watching movies in this experiment 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 developed and manufactured by our chemical sensors laboratory [Brady, Diamond & Lau, 2005], the electrical conductivity of which changes when it is compressed. The polyurethane foam is soft and flexible and it is coated with polypyrrole which is a conducting material, and therefore it has a particular resistance when connected into a circuit and electricity flows through it. If the foam is compressed (the polypyrrole doesn't interfere with the soft/flexible characteristics of the foam), then more polypyrrole is in contact with each other and also the molecule chains of polypyrrole can line up closer to one another, enabling the electricity to flow through more easily - which means the resistance of the foam decreases. Then when the foam is relaxed the resistance returns to the original value.

Movements such as sitting back, moving forward or shifting from side to side in the chair result in pressure changes on the backrest causing the polyurethane foam to compress and relax accordingly and the resulting changes in the electrical current passing through the foam are transmitted to and recorded on a nearby PC. This data allows for the fine-grained monitoring of subjects' movements.

3.2 THE CDV*PLEX* EXPERIMENT

The aim of the CDV*Plex* movie experiment was to provide a ground-truth of film viewers' physiological responses over multiple full-length feature films, by gathering biometric data from viewers using the multiple sensors described earlier in a controlled cinema-like environment, which became known as "The CDV*Plex*" [Rothwell et al, 2006]. This section describes the laboratory set-up and procedure, as well as details of the feature films used and the volunteer human subjects.

3.2.1 The Cinema

In order to obtain as natural and therefore as accurate results as possible, our aim was to encourage viewers to relax and to essentially forget that they were participating in a laboratory experiment. Therefore it was decided to recreate an environment that was as close to a true cinematic experience as possible. Thus, the CDV*Plex* was constructed in an air-conditioned windowless room in which a Dolby 5.1 surround sound speaker system, a DVD player and large-screen digital projector were installed. It seated up to four people at a time and the film was rear-projected onto a large screen at one end of the room. Each of the chairs was outfitted as a Smart Chair (described in the previous section).

3.2.2 The Movies

In total, fifty-one full-length feature films were shown over a ten-week period. These were chosen to encompass a range across seven decades, from 1942 to 2004, and ten movie genres, shown in Table 3.2(a) below. The reader should refer to Appendix B for a full listing of the feature films selected for this experiment.

Movie Genre	**Sample Movies**
Action/Adventure	*Indiana Jones, Lawrence of Arabia*
Animation	*Finding Nemo, Shrek*
Comedy	*Who Framed Roger Rabbit?, Toy Story*
Documentary	*Roger and Me*
Drama	*The Godfather, Pulp Fiction*
Foreign/World Cinema	*Das Boot, Il Postino*
Horror/Suspense	*Alien, The Changeling*
Romance	*Casablanca, The English Patient*
Science Fiction/Fantasy	*Star Wars, Harry Potter*
Thriller/Mystery	*Minority Report, Mean Streets, Leon*

Table 3.2(a): A selection of the films shown in the CDV*Plex*

3.2.3 The Viewers

The participants in this experiment were forty-three volunteers from across the university's population of staff and graduate students, of which eight were female. Film showings were advertised daily using posters and email bulletins. An online booking system and website were provided where interested parties could view the film schedule and the availability of places, and reserve a seat for their chosen film. Participation was limited to at least two and at most four film viewings per person to ensure that we could compare results sets for the same person and yet still maximise participation.

As can be seen from the results of the CDV*Plex* participant surveys below (see Appendix C for a copy of the survey questionnaire), the vast majority of viewers are in their twenties, mostly their early twenties, as for the most part they were made up of graduate students from the university, Figure 3.2(a). In almost half the viewings, participants chose films they hadn't seen before, Figure 3.2(b), and preferences for different aspects were similar across both genders, Figure 3.2(c). Most males typically attended the cinema more or least once a month, with most females attending more frequently – with a large proportion attending once a week, Figure 3.2(d).

In order to ensure participants anonymity, throughout this thesis each viewer will be referred to by the identification code that was assigned to them when they agreed to participate in the CDV*Plex* experiment. For example, P09 or P10-f, where the "-f" denotes that the subject is female.

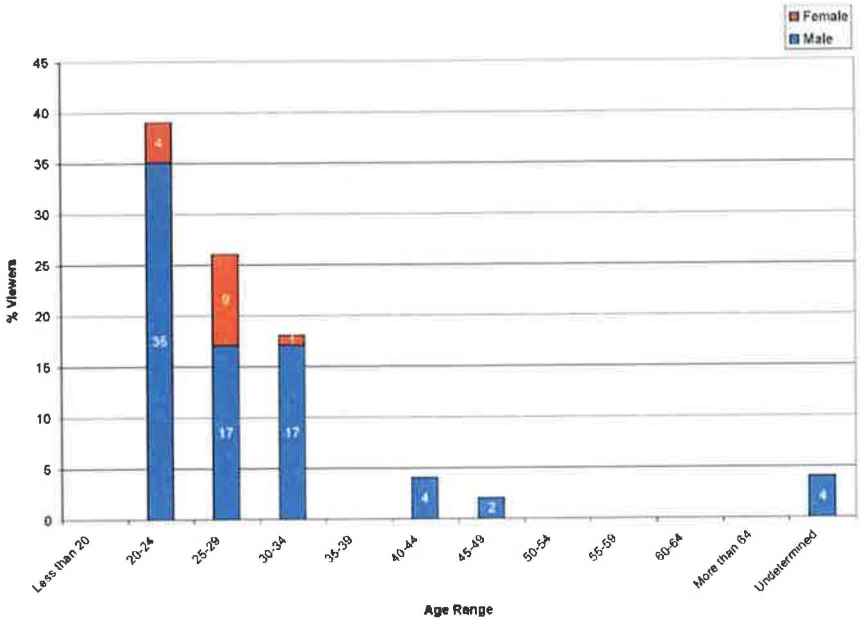


Figure 3.2(a): Age Range of Viewers who Participated in the CDV*Plex*

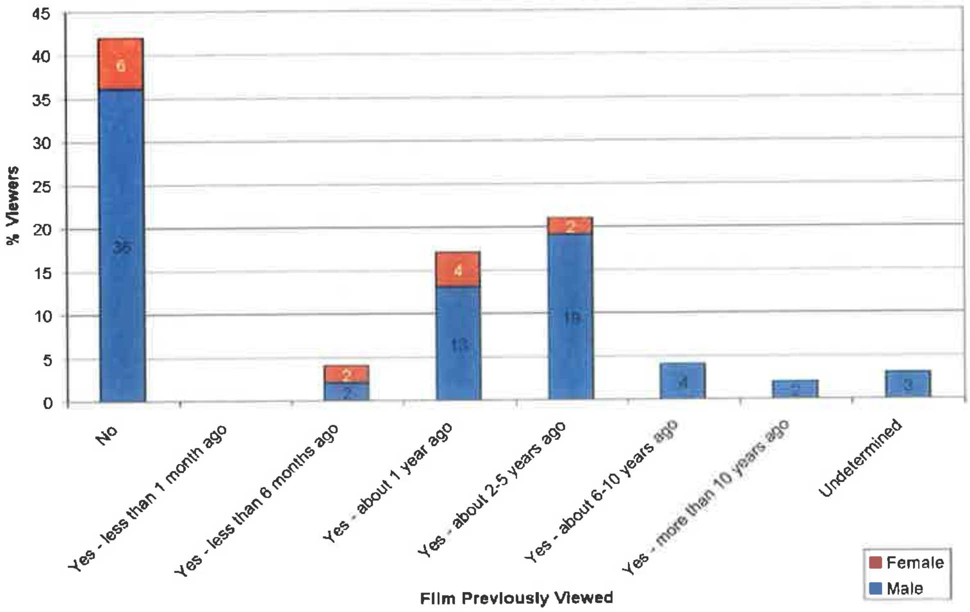


Figure 3.2(b): Viewings of Previously Seen or Unseen Films by Volunteers in the CDV*Plex*

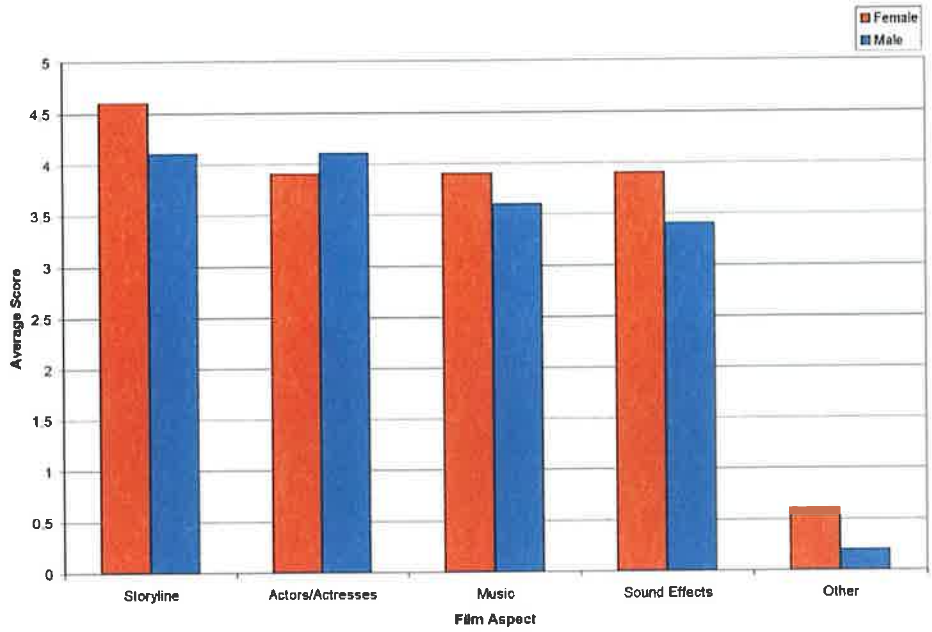


Figure 3.2(c): Aspects of the Films in the CDV*Plex* as Favoured by the Viewers

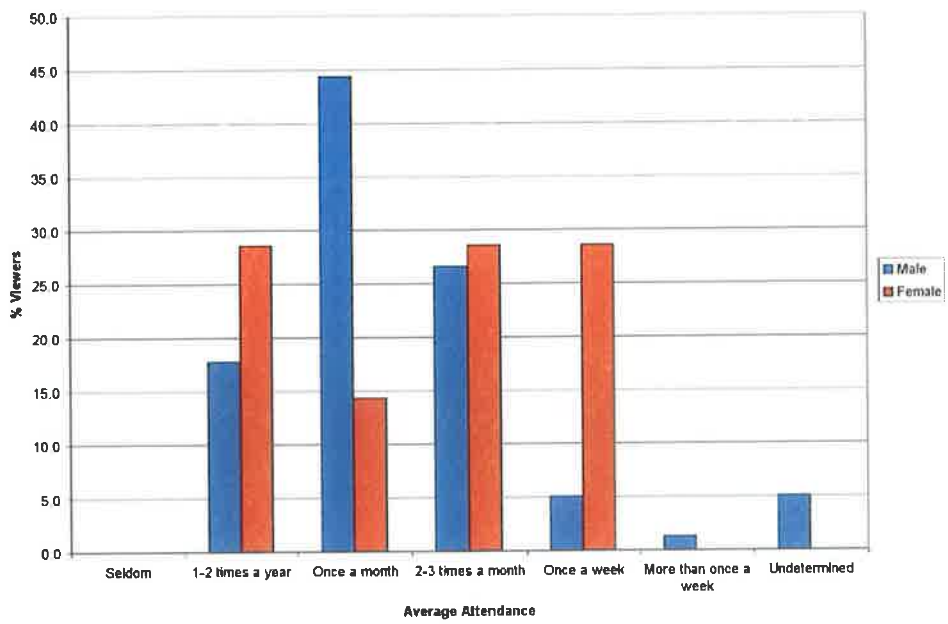


Figure 3.2(d): Typical Cinema Attendance of CDV*Plex* Viewers

3.2.4 *The Procedure*

Once a volunteer had reserved his/her place for a film showing they were contacted with a confirmation of their booking and an information sheet detailing the procedure, requirements and rules of CDV*Plex* participation (please refer to Appendix C for details).

On the day of viewing, each participant was contacted at 11am and issued with the two wearable sensor devices described in the previous section (a heart-rate monitor and a BodyMedia armband). They were required to wear these devices continuously for approximately five to six hours, or until after the film had ended.

The benefits of this extra time for data sampling were two-fold. It allowed the viewers to become accustomed to and comfortable with the devices long before the film started and in their own environments. Secondly, the data recorded in the hours leading up to the film showing established a baseline of biometric responses for that person. To that end, participants were asked to refrain from engaging in strenuous physical activities such as exercise (for example, jogging or swimming) during this period, as such activities would skew the readings.

On entering the CDV*Plex* at 2pm, viewers were seated and asked to complete a short questionnaire detailing some personal information such as age and gender, their cinema attendance and preferred film genres, and also whether or not they had previously seen the film they were about to watch (see Section 3.2.3 above). This also allowed time for the Smart Chairs to be calibrated and baselines established for each viewer. The procedure for time-stamping their devices and viewing the film was explained and at 2:30pm the room was darkened, and the film was started with each viewer synchronously time-stamping their heart-rate monitor and armband device.

At the end of the film, once the titles and credits had finished rolling, the timestamps were again synchronised. Participants were then asked to complete another brief questionnaire on which aspects or components of the film they most enjoyed (plot, music, etc) and were then asked to list their top five highlights from the film and how each made them feel. The devices were then removed and the biometric data downloaded and stored.

3.3 BIOMETRIC DATA

In all, thirty-one full-length feature films shown over ten weeks to forty-three volunteer participants resulted in over 500 hours of recorded biometric observations, half as baseline observations of people going about their daily lives without any deliberate strenuous activity and half involving people watching movies.

For the purposes of this work, only six films were selected for further analysis, the details of and justification for which are given later in Chapter 5. The three streams of data for each of these six films were recorded and stored as separate independent data files in a variety of formats, mostly plain text files. The relevant numerical data was extracted from these files and in order to analyse the biometric data in terms of the film viewing, the data was split using the timestamp information. The data then underwent some simple processing to aggregate the various streams into time series of 5s intervals and also 60s intervals in order to render it into a form which could then be processed to detect events of interest in the data which could indicate significant physiological reactions in the subjects under investigation. The detection process for each type of biometric measurement, GSR, heart-rate and motion, is discussed below.

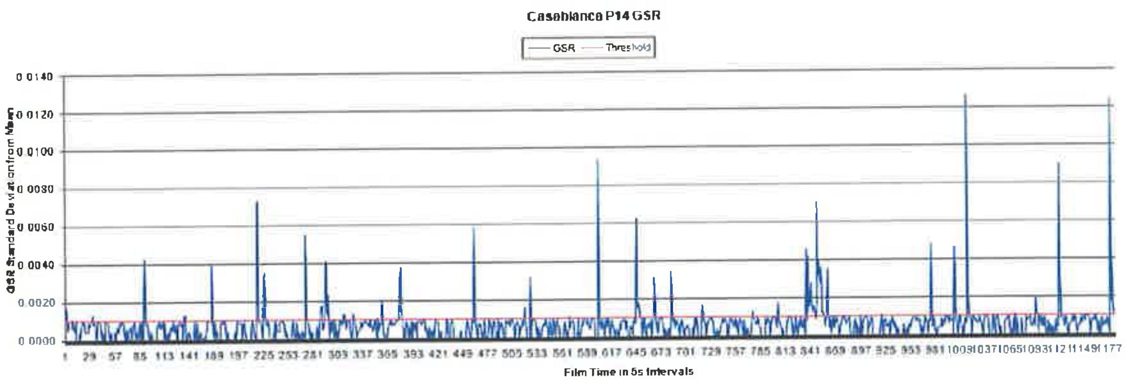
Unfortunately in some cases the biometric data became corrupted or the devices did not function properly and therefore for some viewings of certain films not all three data types are available for analysis. Fortunately no GSR data was affected, however the heart-rate monitor readings for P23-f in *Finding Nemo* and P10-f in *Pirates of the Caribbean* were completely corrupted, as were the first 25 minutes of P04 in *Pirates of the Caribbean*, due to some interference issues with the wireless receiver in the heart-rate device. Unfortunately no motion sensor data was recorded for *Finding Nemo* for subjects P10-f, P23-f, P28-f and P35-f due to a malfunction with the Smart Chair system on the occasion of that viewing. However this still left us with more than enough data to analyse with forty-two data streams from sixteen viewings.

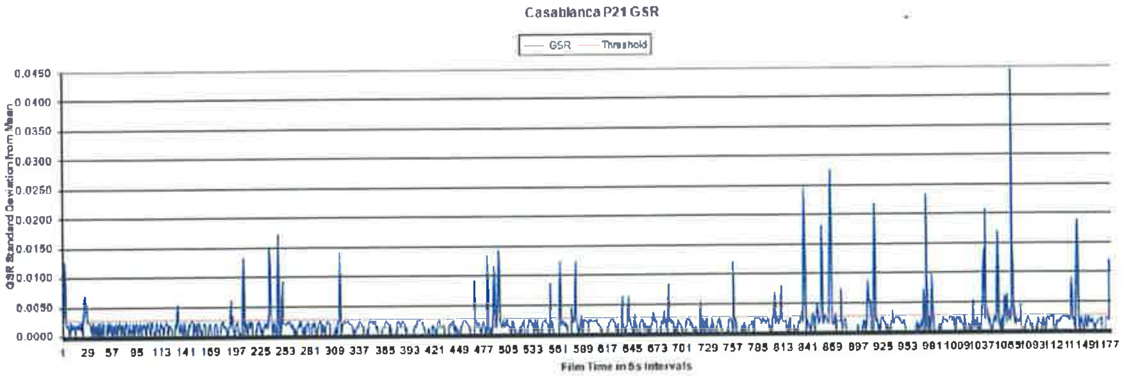
3.3.1 GSR Peak Detection

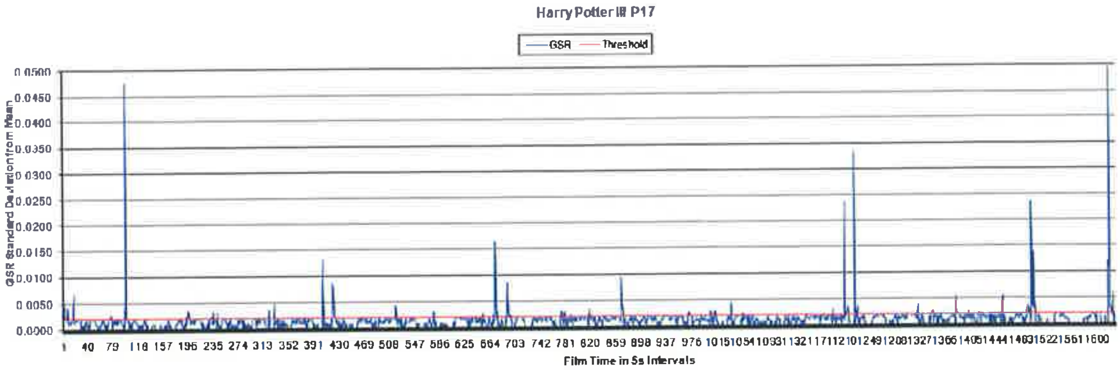
In order to detect peaks in the GSR data for each viewing we first calculated the standard deviation from the mean for each GSR data series which revealed any substantial changes in the readings. Based on this data we identified a threshold of significance for each individual viewing and applying these thresholds to the data allowed us to indentify peaks in the GSR readings. These thresholds and the number of peaks that were present in each set of GSR data are detailed below in Table 3.3(a) and the results are illustrated below in Figure 3.3(a). Once the GSR data had been processed for the occurrence of physiological peaks, the magnitude, or relative size, of each peak was calculated as a percentage of the maximum and the results were plotted below in Figure 3.3(b).

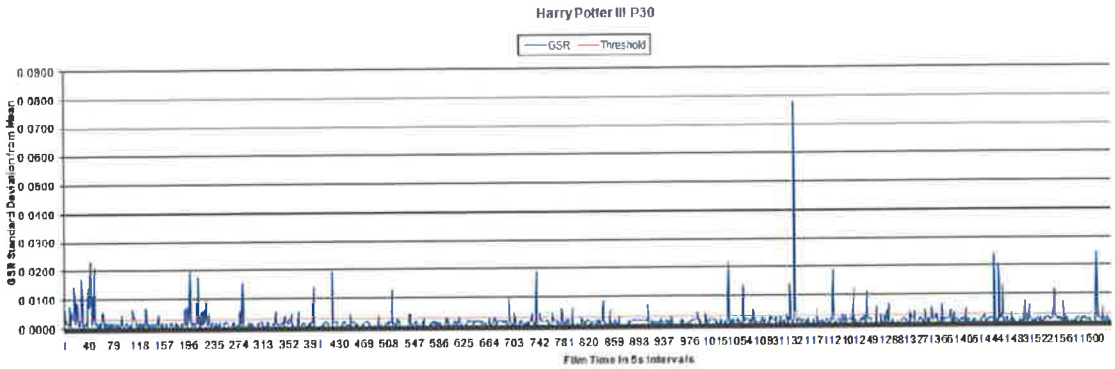
Film	Viewer	GSR Std. Deviation Threshold	No. Peaks	No. Peaks >25% Magnitude	No. Peaks >50% Magnitude
Casablanca	P14	0.0010	83	21	7
	P21	0.0025	93	20	3
Finding Nemo	P35-F	0.0020	59	11	5
	P28-F	0.0050	253	0	0
	P23-F	0.0003	218	6	2
	P10-F	0.0020	37	5	2
Harry Potter III	P30	0.0030	140	4	1
	P17	0.0020	92	7	3
Indiana Jones III	P09	0.0035	150	4	3
	P37	0.0020	93	9	7
Mean Streets	P12	0.0040	115	13	3
	P18	0.0050	84	10	3
	P36	0.0035	131	34	8
Pirates of the Caribbean	P04	0.0008	78	9	5
	P10-F	0.0009	132	16	7
	P11	0.0030	137	10	5

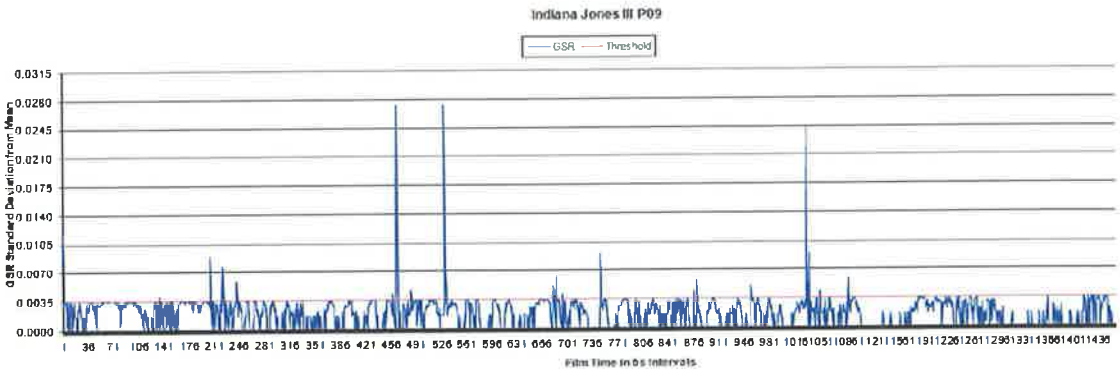
Table 3.3(a): Thresholds and Peak Detection in GSR Data per Viewing

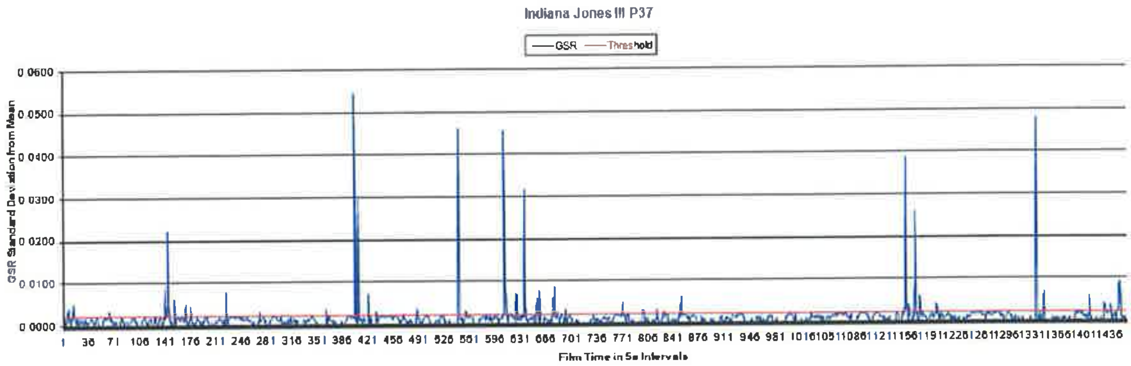


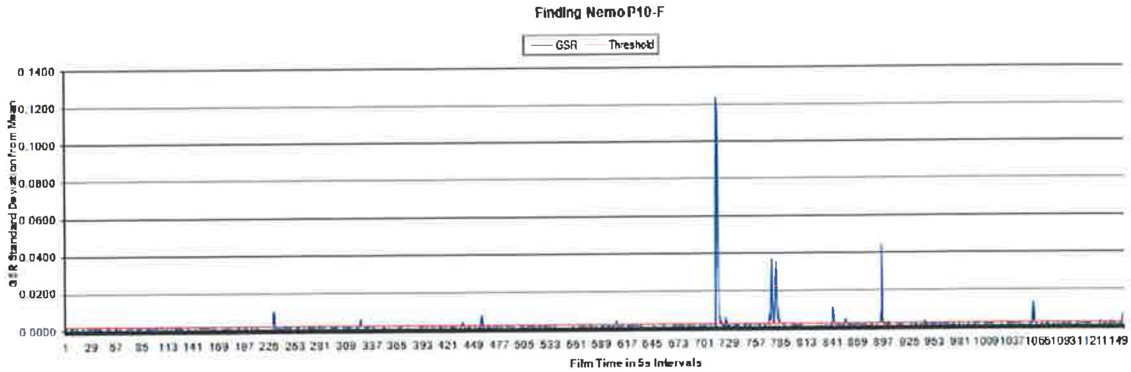

Casablanca P21 GSR
GSR
Threshold
GSR Standard Deviation from Mean
0.0450
0.0400
0.0350
0.0300
0.0250
0.0200
0.0150
0.0100
0.0050
0.0000
Film Time in 5s Intervals

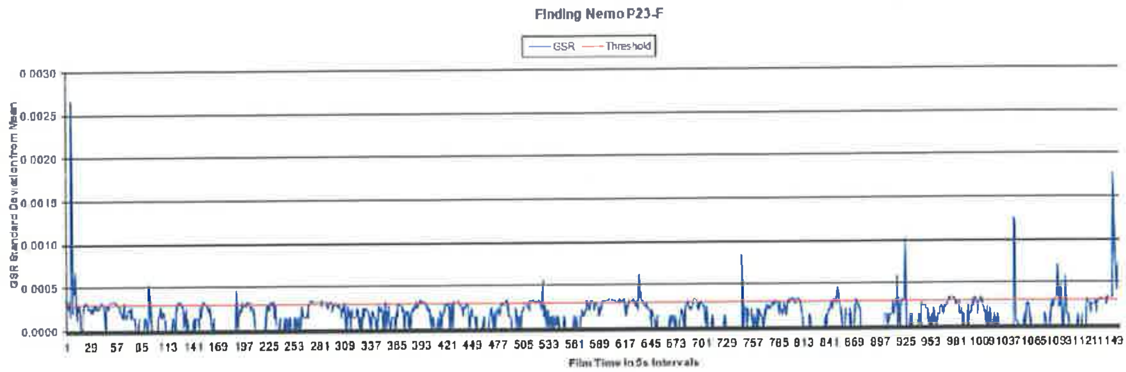

Harry Potter III P17
GSR
Threshold
GSR Standard Deviation from Mean
0.0500
0.0450
0.0400
0.0350
0.0300
0.0250
0.0200
0.0150
0.0100
0.0050
0.0000
Film Time in 5s Intervals

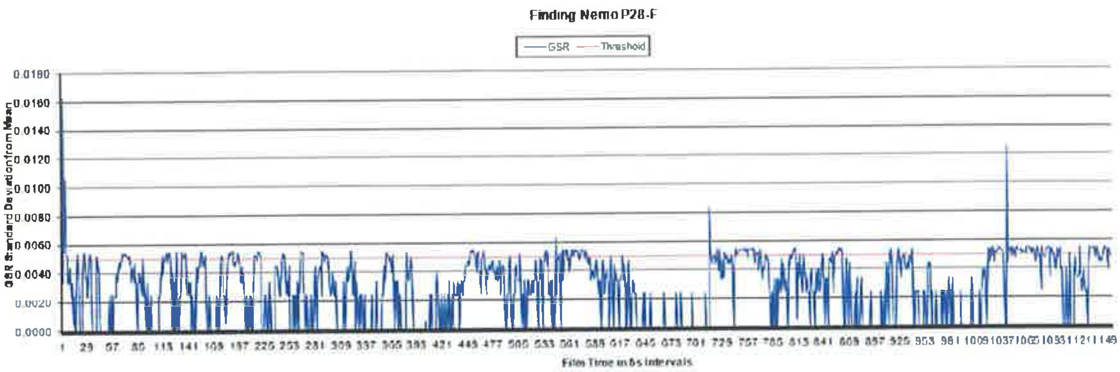

Harry Potter III P30
GSR
Threshold
GSR Standard Deviation from Mean
0.0900
0.0800
0.0700
0.0600
0.0500
0.0400
0.0300
0.0200
0.0100
0.0000
Film Time in 5s Intervals

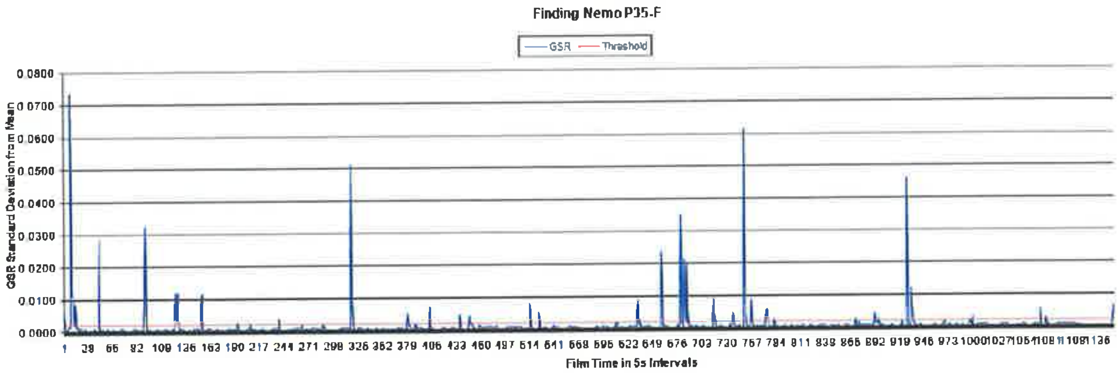

Indiana Jones III P09
GSR
Threshold
GSR Standard Deviation from Mean
0.0315
0.0280
0.0245
0.0210
0.0175
0.0140
0.0105
0.0070
0.0035
0.0000
Film Time in 5s Intervals

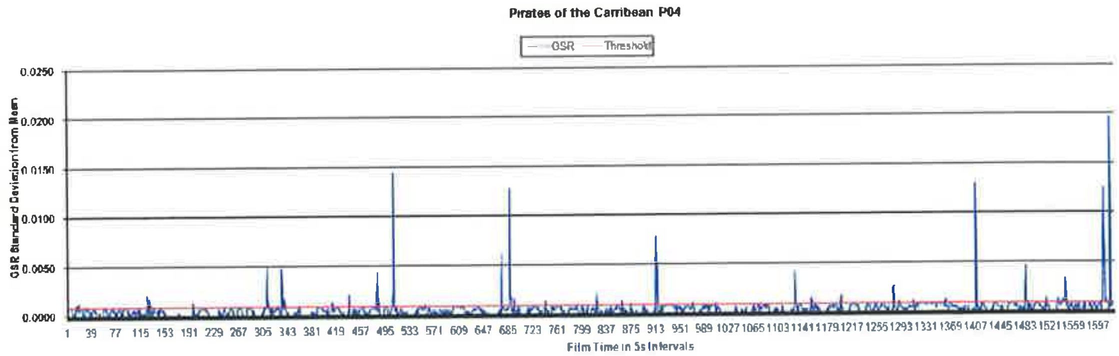

Indiana Jones III P37
GSR
Threshold
GSR Standard Deviation from Mean
Film Time in 5s Intervals

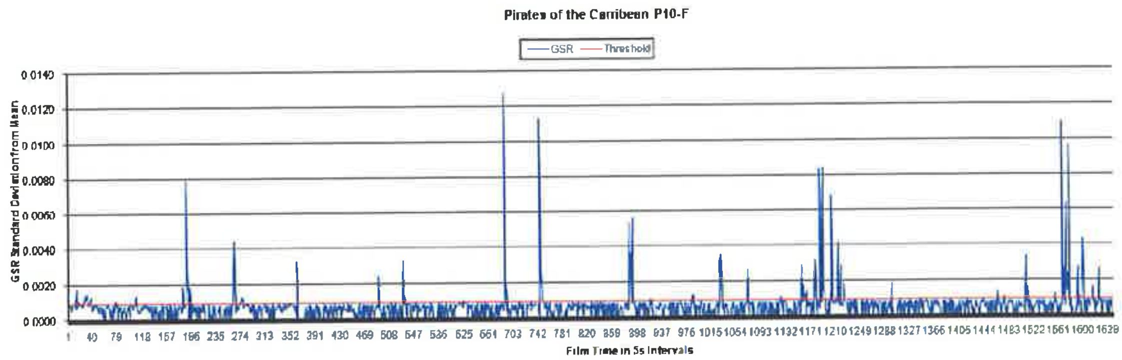

Finding Nemo P10-F
GSR
Threshold
GSR Standard Deviation from Mean
Film Time in 5s Intervals

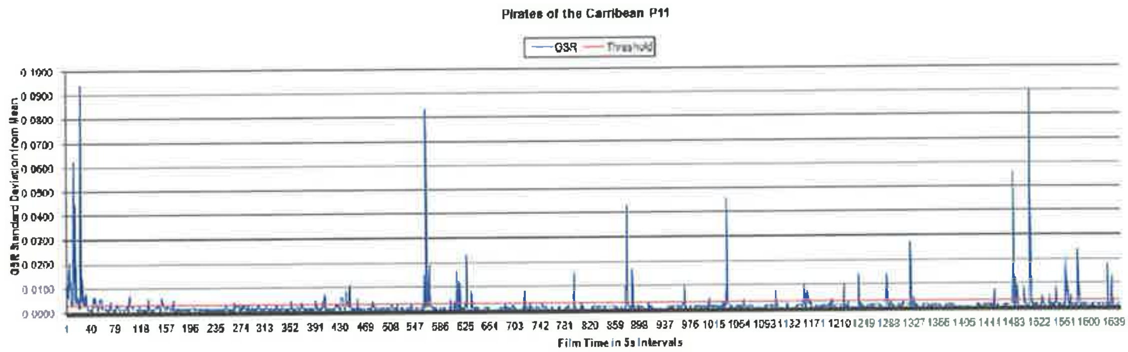

Finding Nemo P23-F
GSR
Threshold
GSR Standard Deviation from Mean
Film Time in 5s Intervals


Finding Nemo P28-F
GSR
Threshold
GSR Standard Deviation from Mean
Film Time in 5s Intervals


Finding Nemo P35-F
GSR
Threshold
GSR Standard Deviation from Mean
0.0000
0.0800
Film Time in 5s Intervals


Pirates of the Carribean P04
GSR
Threshold
GSR Standard Deviation from Mean
0.0000
0.0250
Film Time in 5s Intervals


Pirates of the Carribean P10-F
GSR
Threshold
GSR Standard Deviation from Mean
0.0000
0.0140
Film Time in 5s Intervals


Pirates of the Carribean P11
GSR
Threshold
GSR Standard Deviation from Mean
0.0000
0.1000
Film Time in 5s Intervals

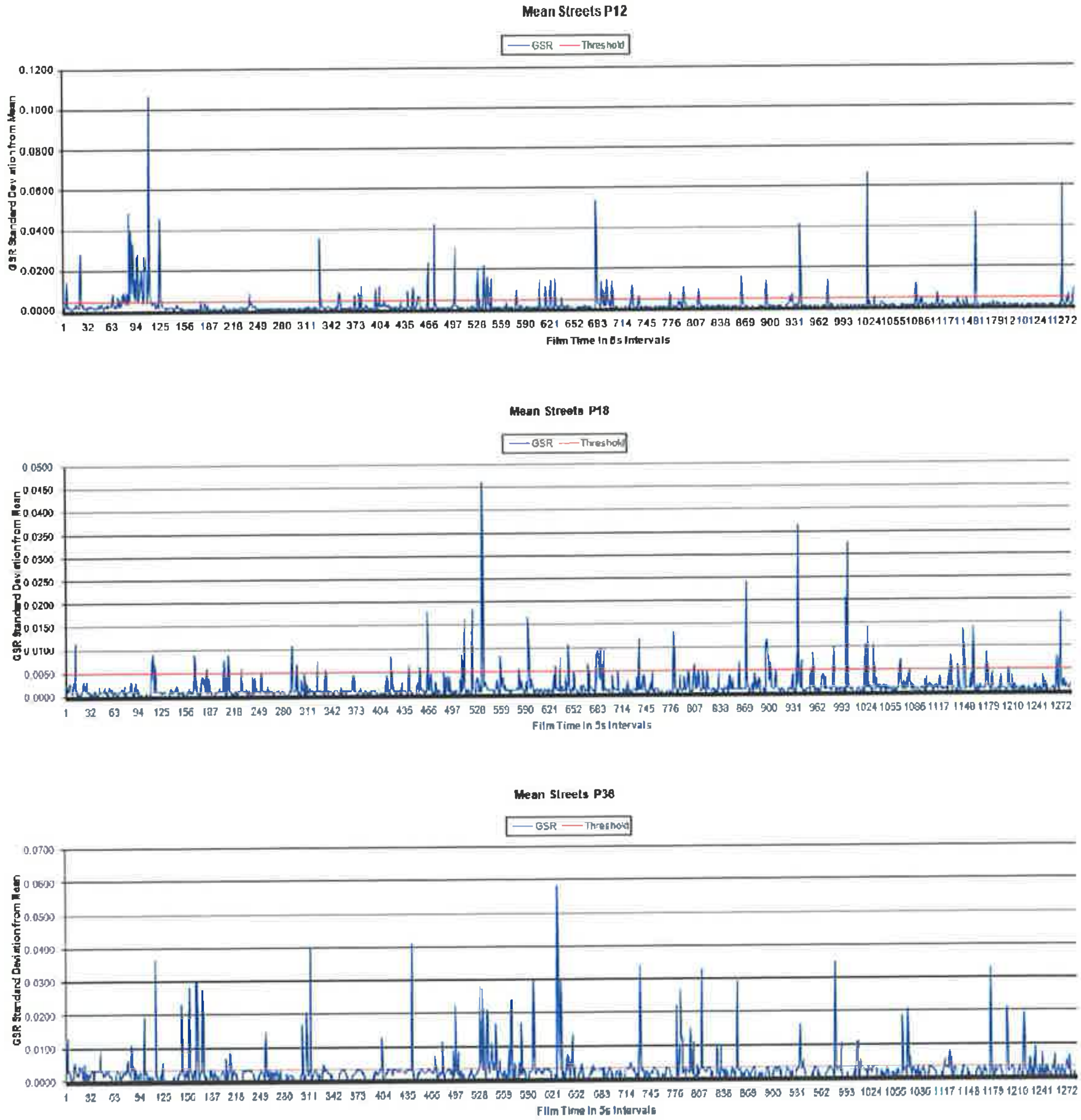
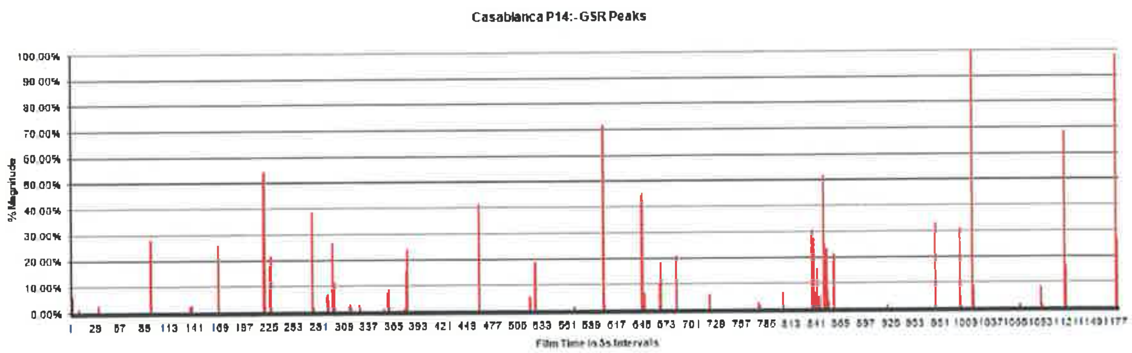
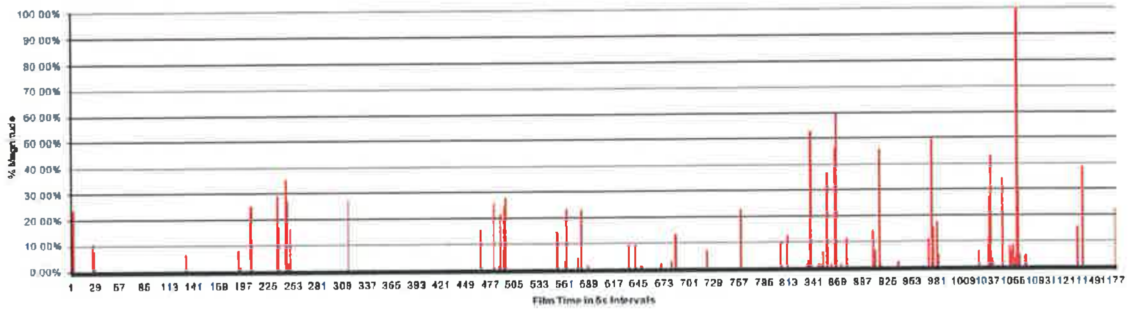


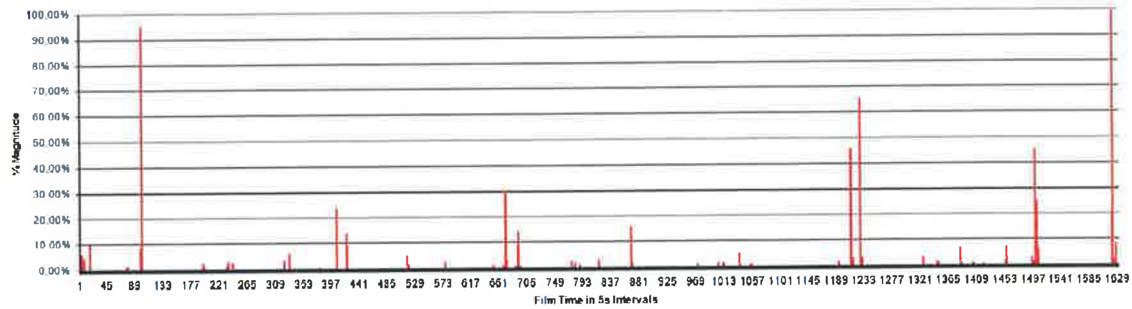
Figure 3.3(a): GSR Peak Detection per Film Viewing



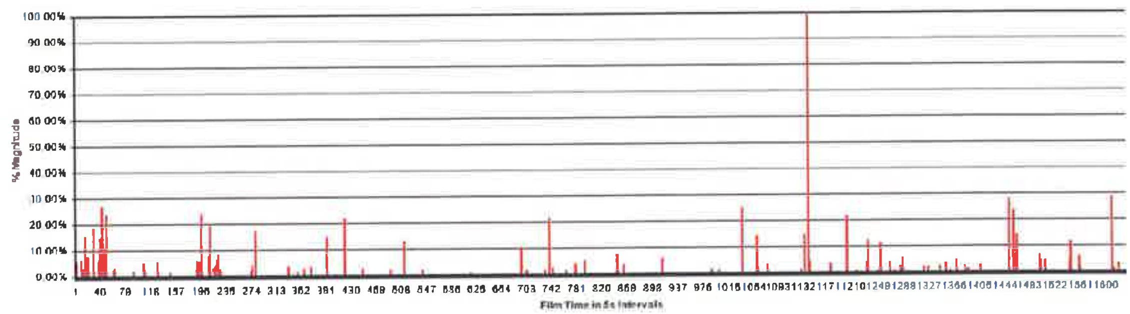
Casablanca P21:- GSR Peaks



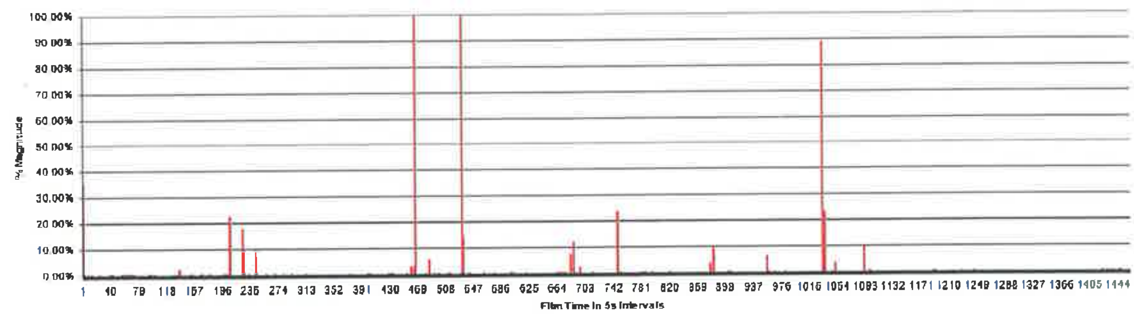
Harry Potter III P17:- GSR Peaks



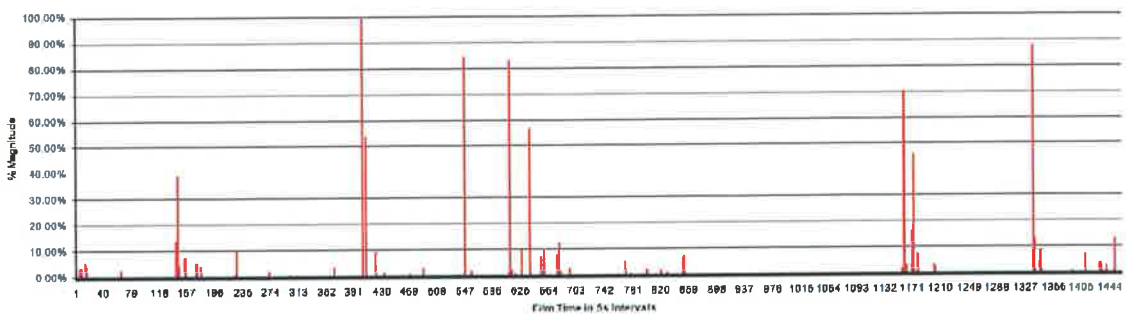
Harry Potter III P30:- GSR Peaks



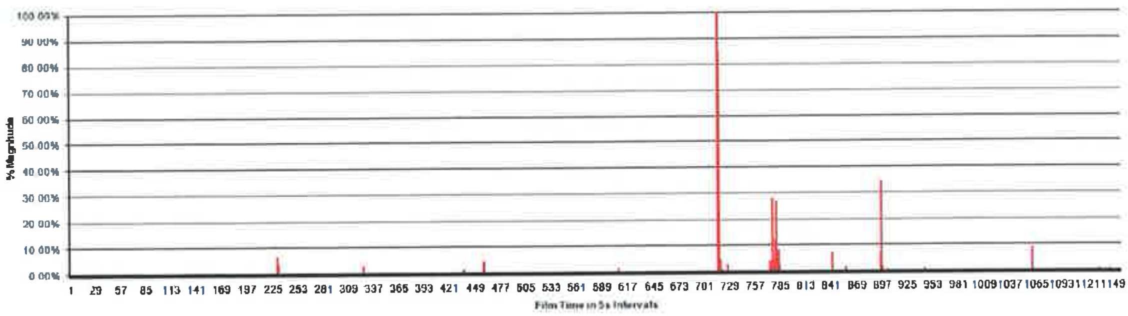
Indiana Jones III P09:- GSR Peaks



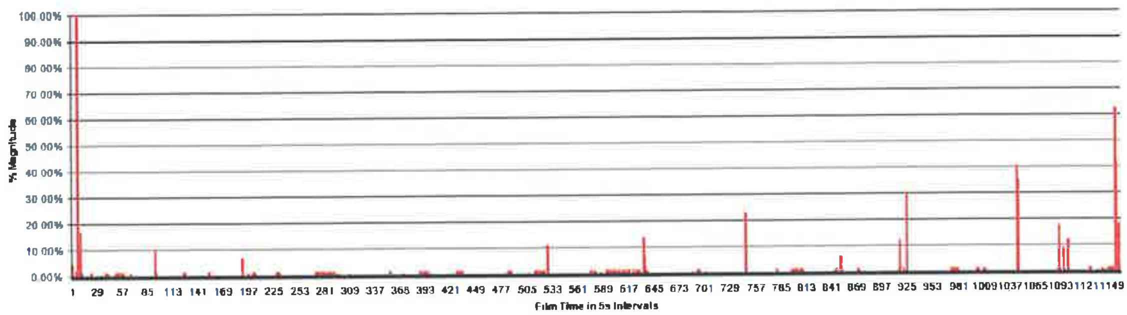
Indiana Jones III P37:- GSR Peaks



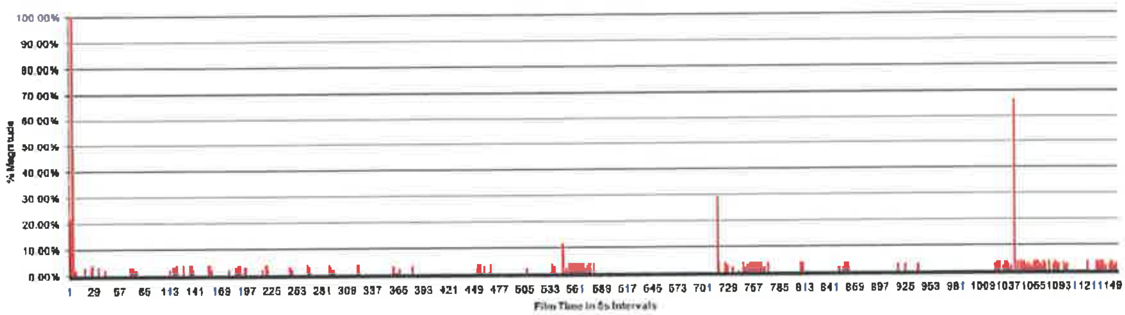
Finding Nemo P10-F:- GSR Peaks

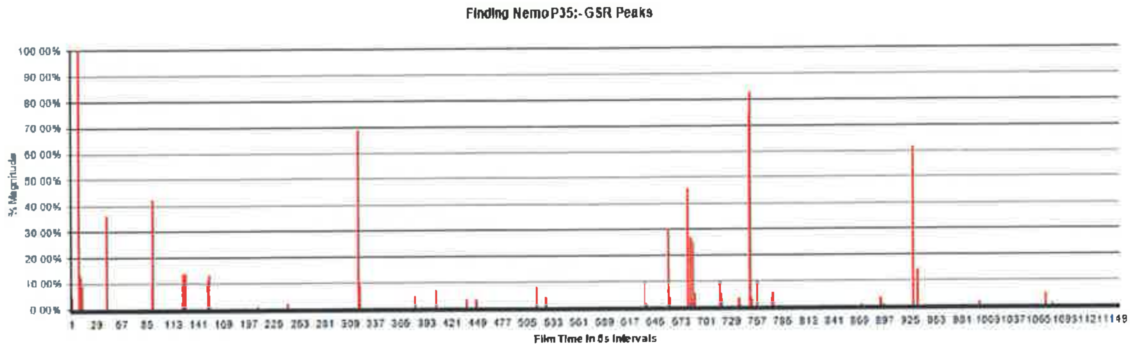


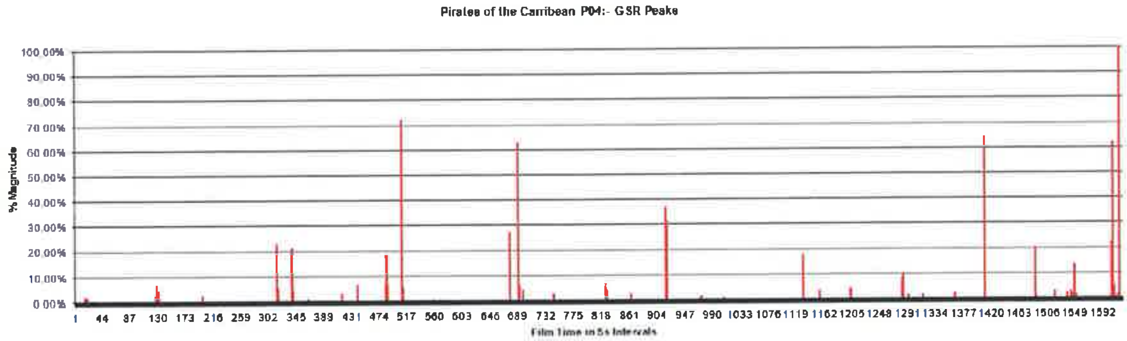
Finding Nemo P23-F:- GSR Peaks

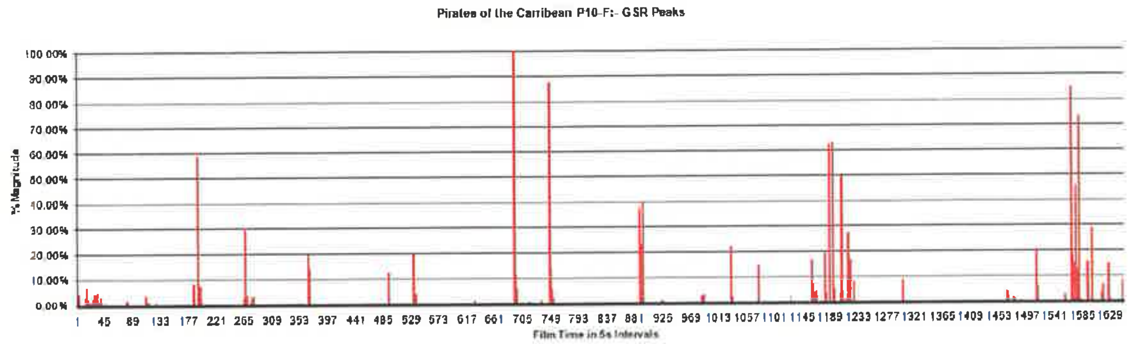


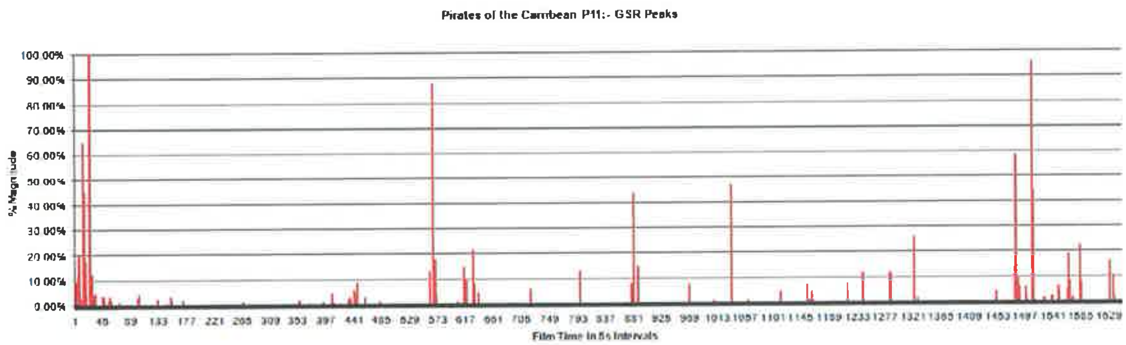
Finding Nemo P28-F:- GSR Peaks




Finding Nemo P35:- GSR Peaks
% Magnitude
Film Time in 5s Intervals


Pirates of the Carribean P04:- GSR Peaks
% Magnitude
Film Time in 5s Intervals


Pirates of the Carribean P10-F:- GSR Peaks
% Magnitude
Film Time in 5s Intervals


Pirates of the Carribean P11:- GSR Peaks
% Magnitude
Film Time in 5s Intervals

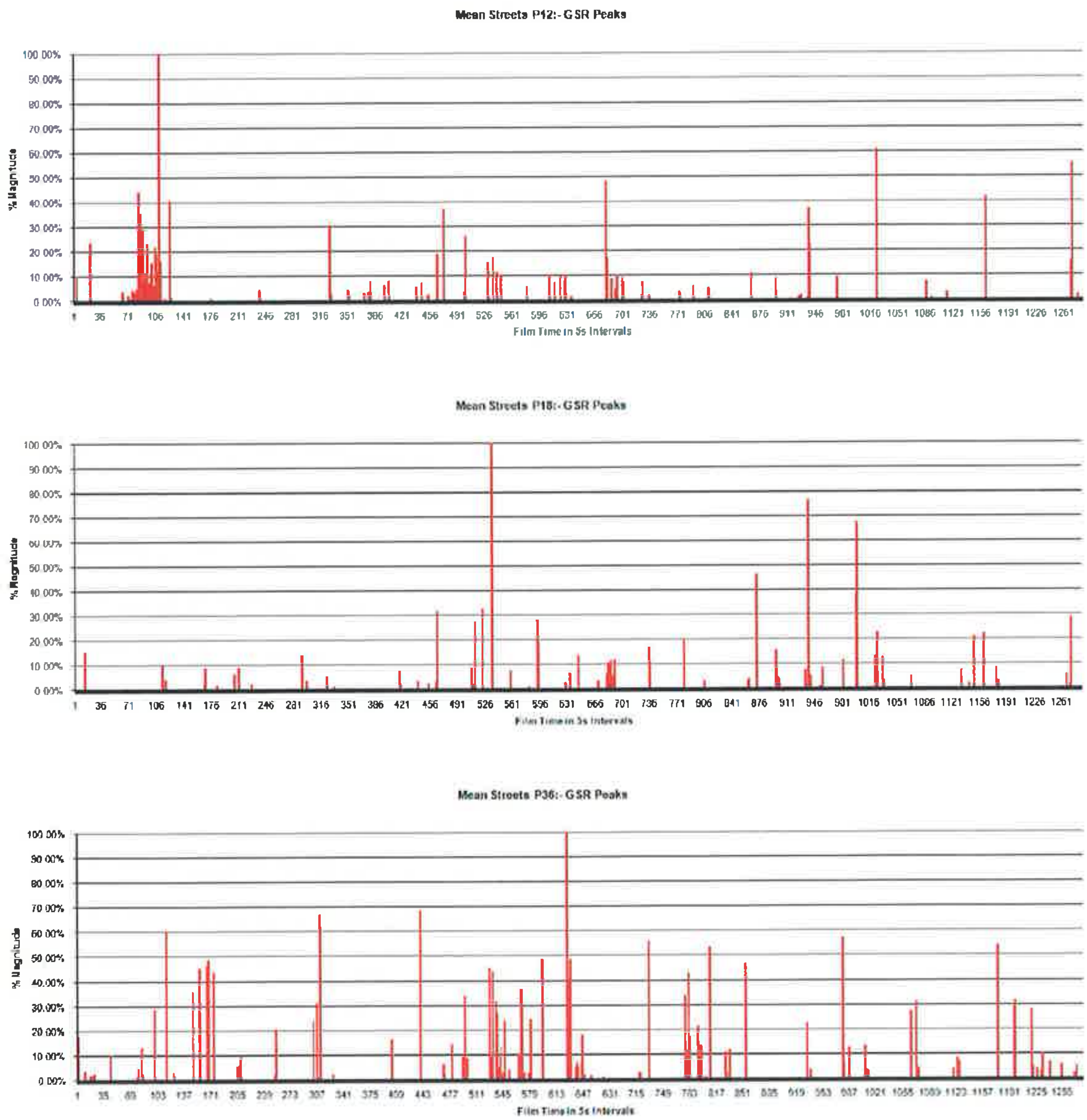


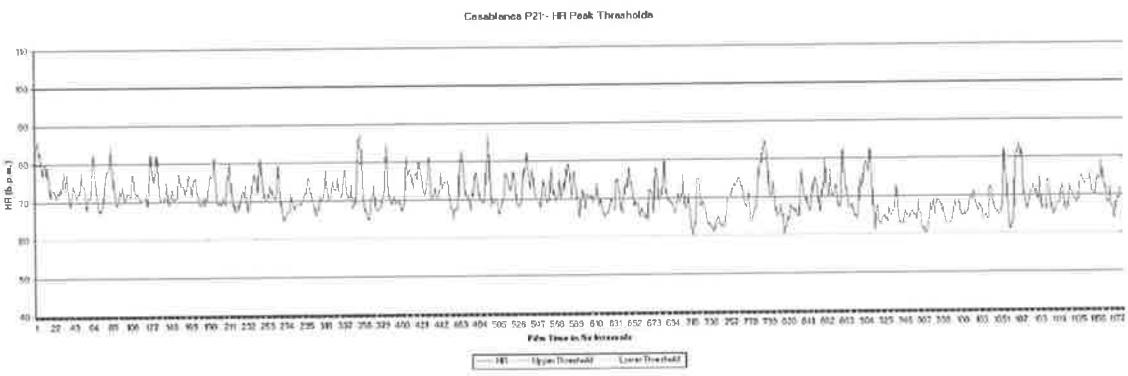
Figure 3.3(b): GSR Peaks Relative Magnitude per Viewing

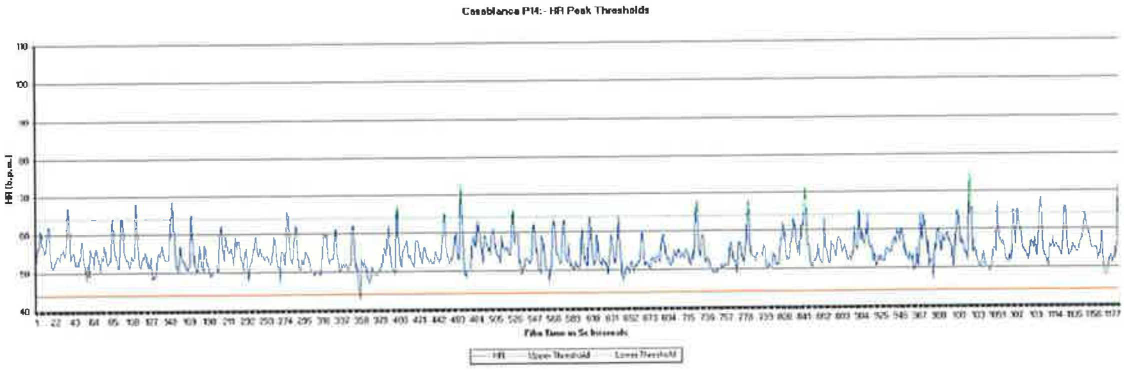
3.3.2 Heart-Rate Peak Detection

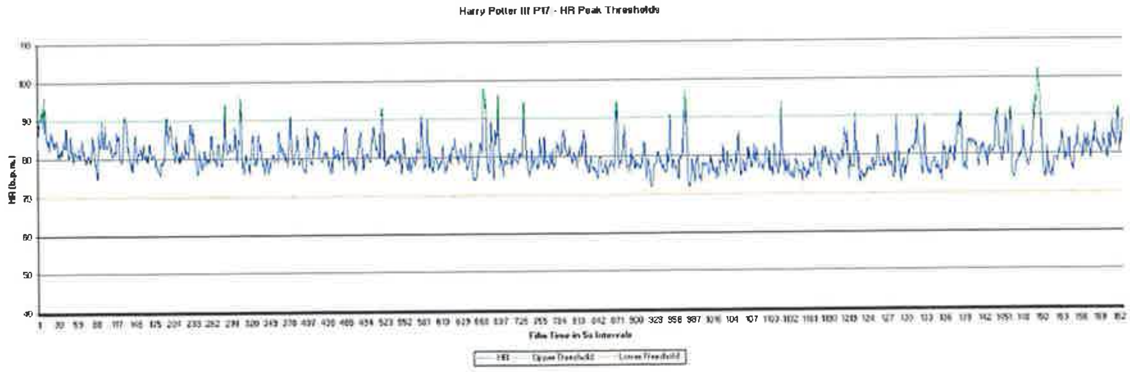
In order to detect peaks in the heart-rate data for each viewing we first calculated the mean baseline for each heart-rate data series and based on this data we identified upper and lower threshold of significance for each individual viewing. Heart-rate is very reactive and fluctuations from the baseline of greater than ± 10 b.p.m.. can be interpreted as a notable event. Applying these thresholds to the data allowed us to indentify peaks in the heart-rate readings. These thresholds and the number of peaks that were present in each set of heart-rate data are detailed below in Table 3.3(b) and the results are illustrated below in Figure 3.3(c). Once the heart-rate data had been processed for the occurrence of physiological peaks, the magnitude, or relative size, of each peak was calculated as a percentage of the maximum and the results were plotted below in Figure 3.3(d).

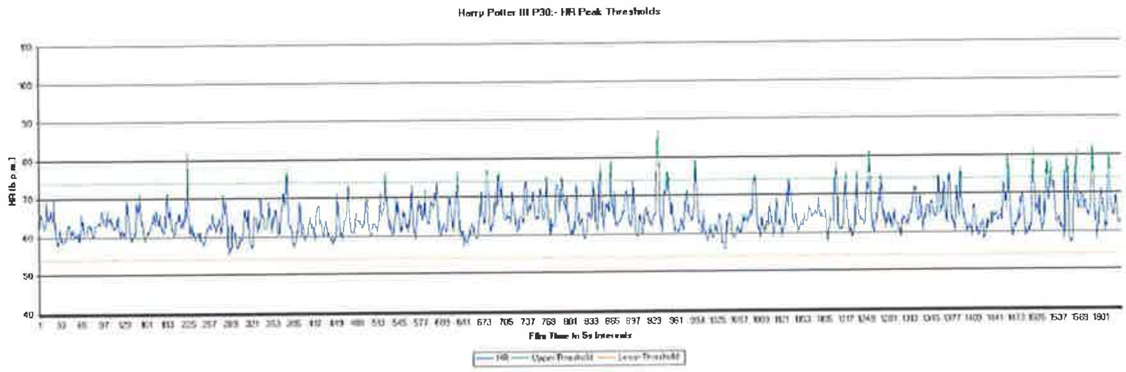
Film	Viewer	Threshold		No. Peaks	No. Peaks >25% Magnitude	No. Peaks >50% Magnitude
		Upper	Lower			
Casablanca	P14	64	44	53	53	53
	P21	80	60	67	67	67
Finding Nemo	P35-F	79	59	33	33	33
	P28-F	82	62	61	61	61
	P23-F	n/a	n/a	n/a	n/a	n/a
	P10-F	69	49	69	69	69
Harry Potter III	P30	74	54	72	72	72
	P17	90	70	62	62	62
Indiana Jones III	P09	83	63	51	51	51
	P37	94	74	75	75	75
Mean Streets	P12	103	83	98	98	98
	P18	73	53	63	63	63
	P36	72	52	116	116	116
Pirates of the Caribbean	P04	83	63	91	91	91
	P10-F	n/a	n/a	n/a	n/a	n/a
	P11	92	72	126	126	126

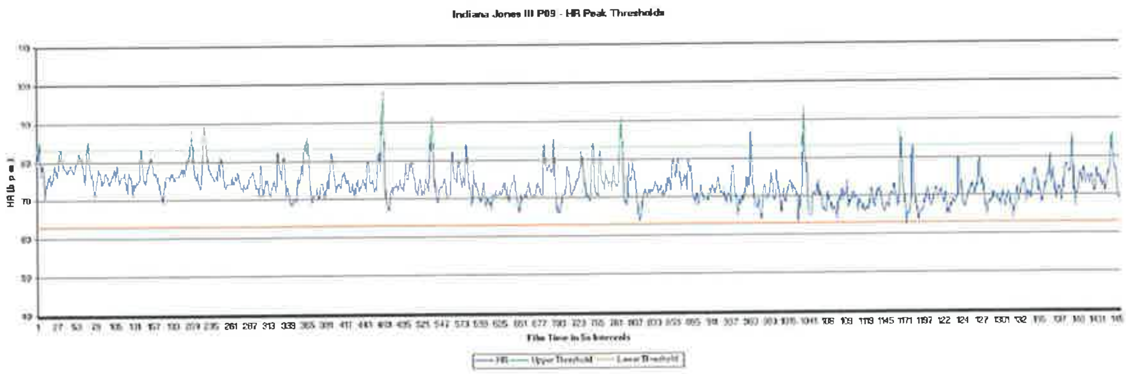
Table 3.3(b): Thresholds and Peak Detection in Heart-Rate Data per Viewing



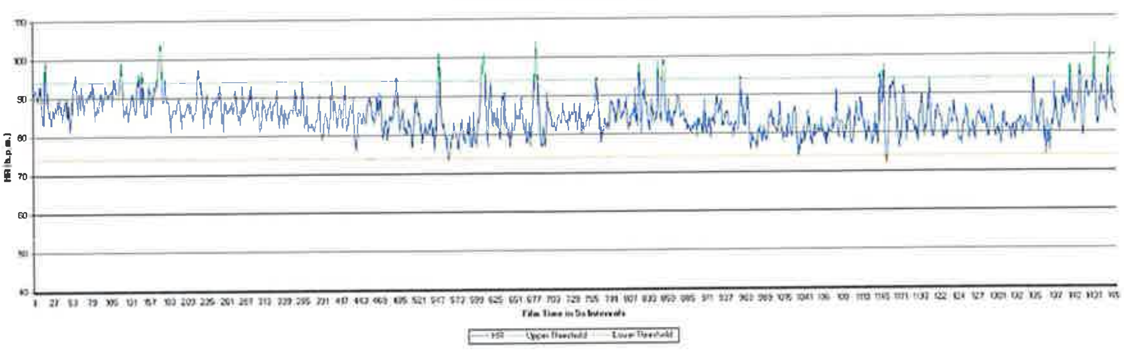

Casablanca P14:- HR Peak Thresholds
HR (b.p.m.)
Film Time in 5s Intervals
HR
Upper Threshold
Lower Threshold


Harry Potter III P17:- HR Peak Thresholds
HR (b.p.m.)
Film Time in 5s Intervals
HR
Upper Threshold
Lower Threshold

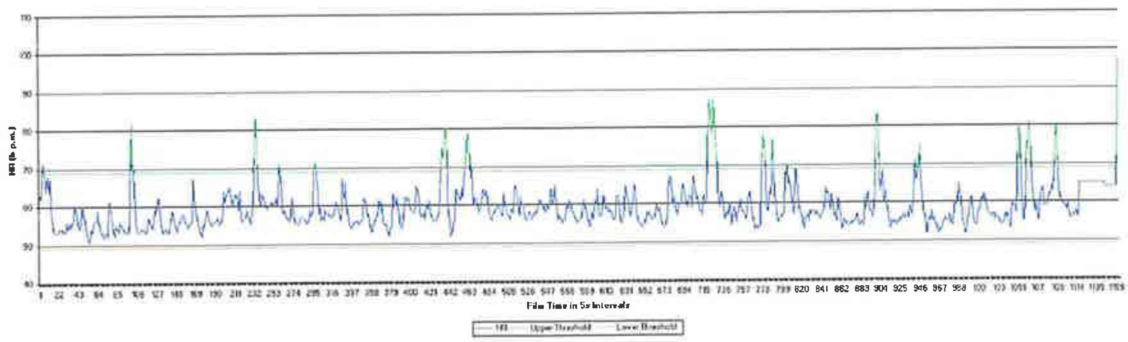

Harry Potter III P30:- HR Peak Thresholds
HR (b.p.m.)
Film Time in 5s Intervals
HR
Upper Threshold
Lower Threshold


Indiana Jones III P09 - HR Peak Thresholds
HR (b.p.m.)
Film Time in 5s Intervals
HR
Upper Threshold
Lower Threshold

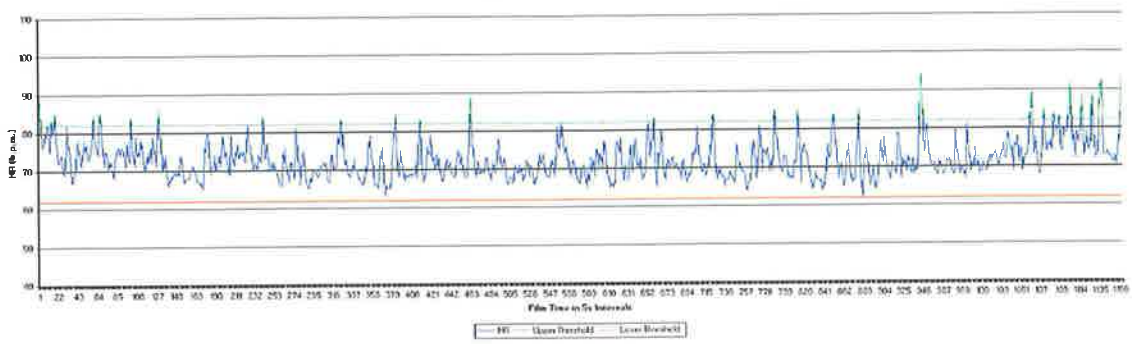
Indiana Jones III P37 - HR Peak Thresholds



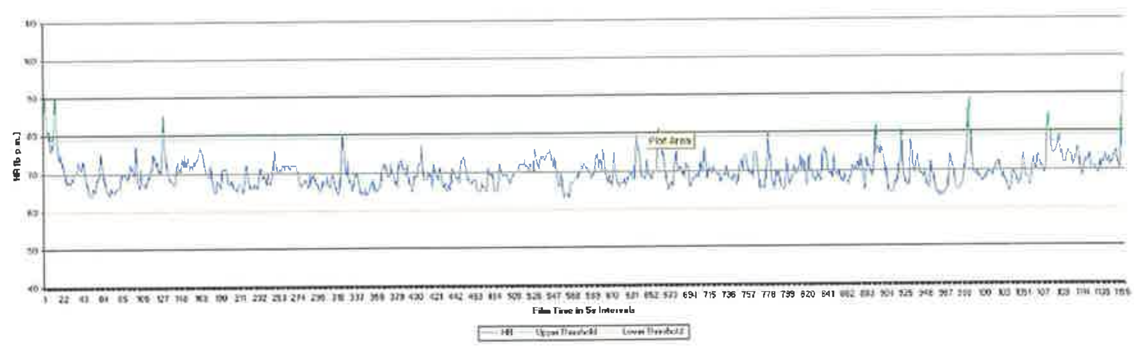
Finding Nemo P10-F.- HR Peak Thresholds

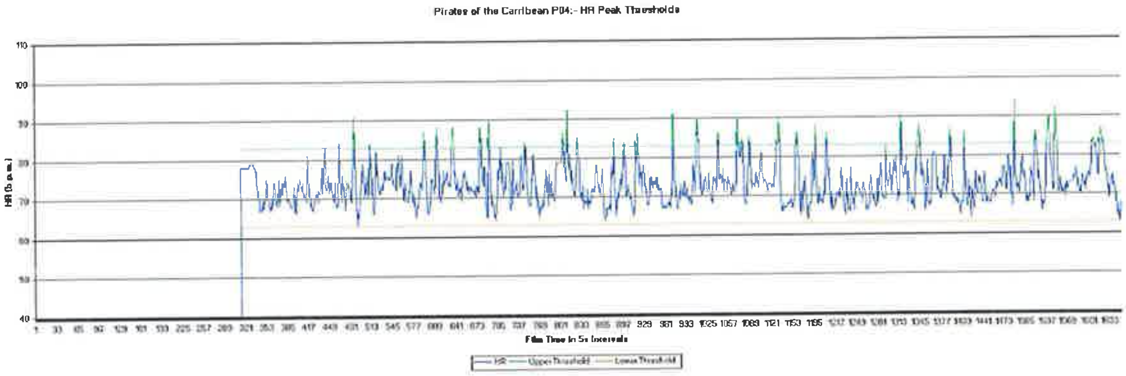


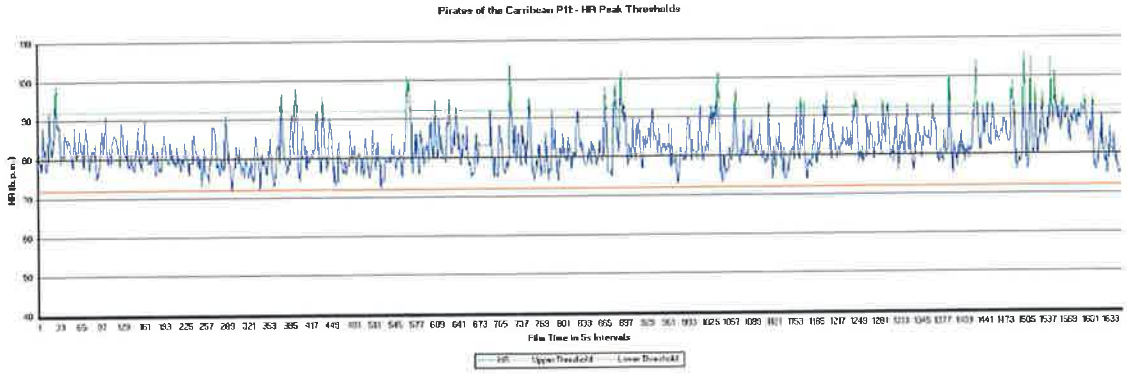
Finding Nemo P28-F.- HR Peak Thresholds

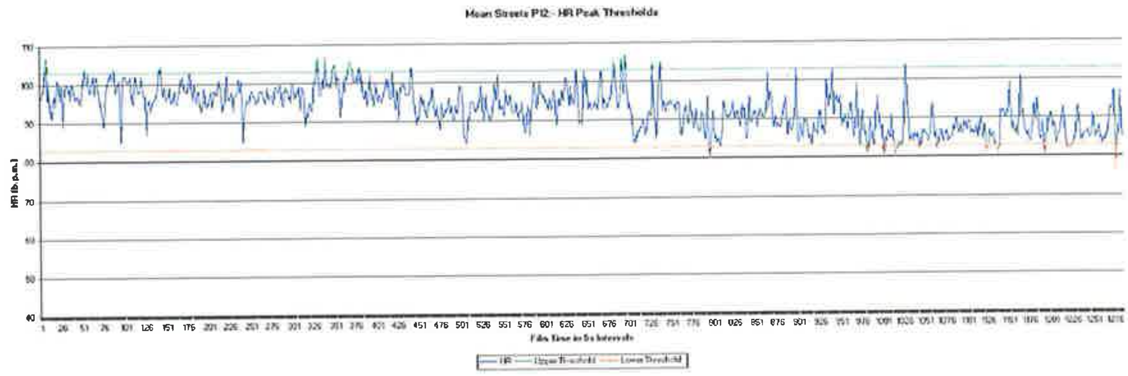


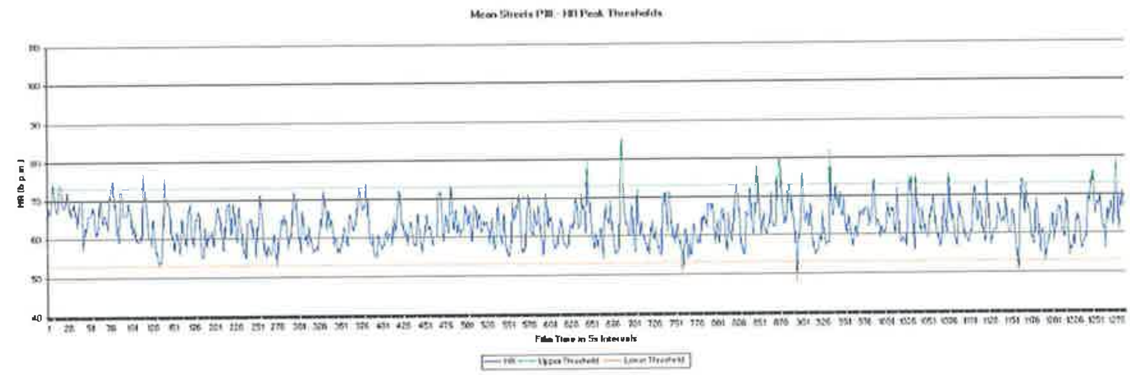
Finding Nemo P35-F.- HR Peak Thresholds




Pirates of the Carribean P04:- HR Peak Thresholds
HR (b.p.m.)
Film Time in 5s Intervals
HR
Upper Threshold
Lower Threshold


Pirates of the Carribean P11 - HR Peak Thresholds
HR (b.p.m.)
Film Time in 5s Intervals
HR
Upper Threshold
Lower Threshold


Mean Streets P12:- HR Peak Thresholds
HR (b.p.m.)
Film Time in 5s Intervals
HR
Upper Threshold
Lower Threshold


HR (b.p.m.)
Film Time in 5s Intervals
HR
Upper Threshold
Lower Threshold

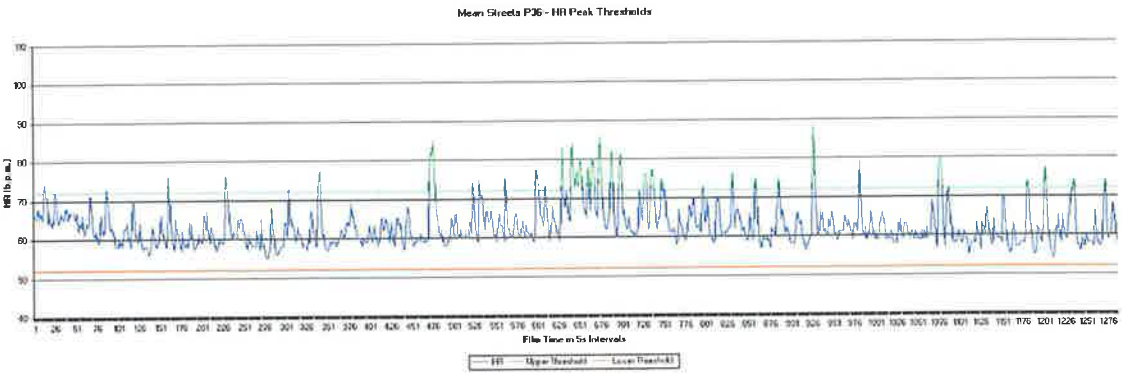
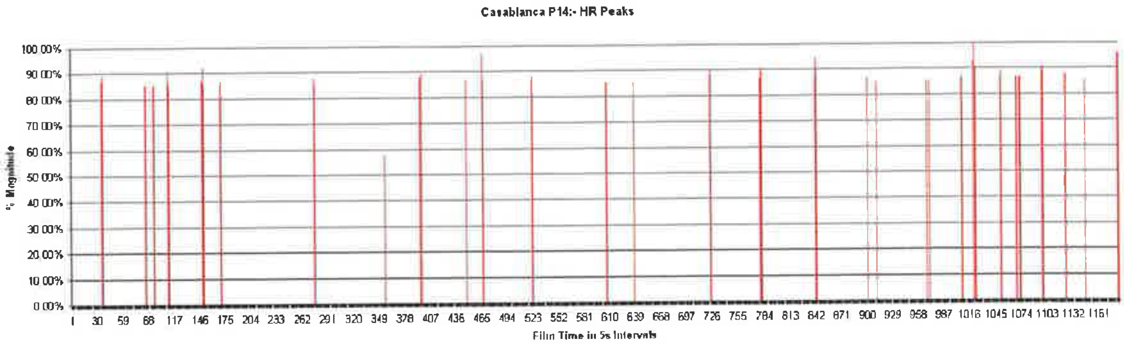
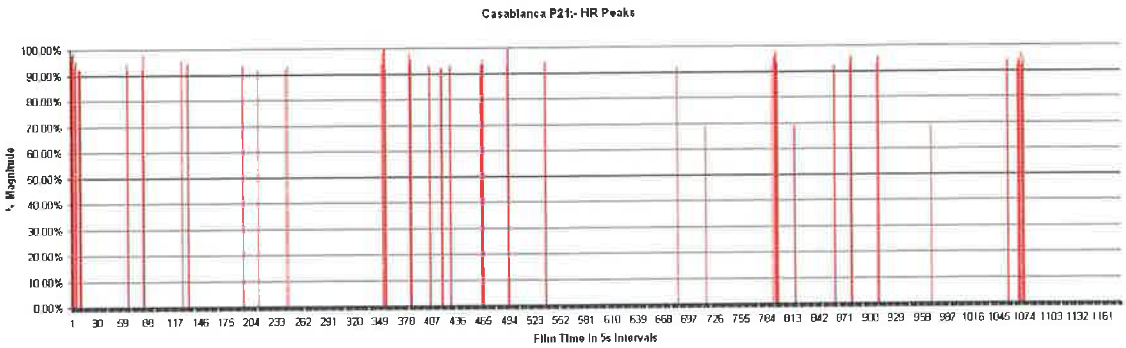
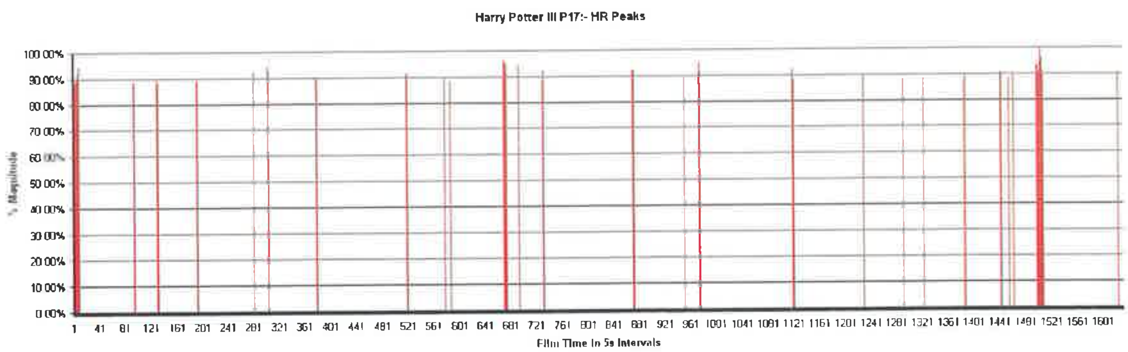


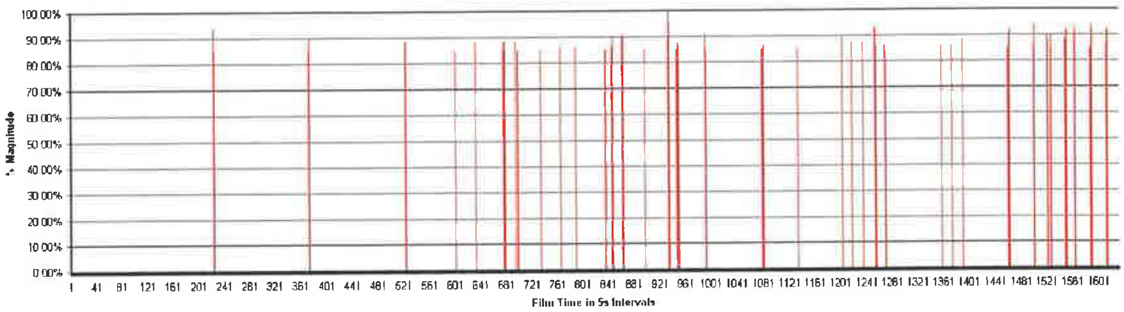
Figure 3.3(c): Heart-Rate Peak Detection per Viewing



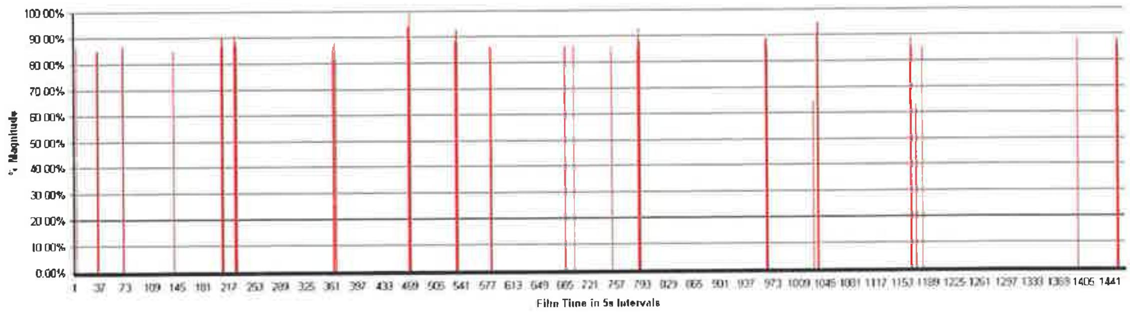




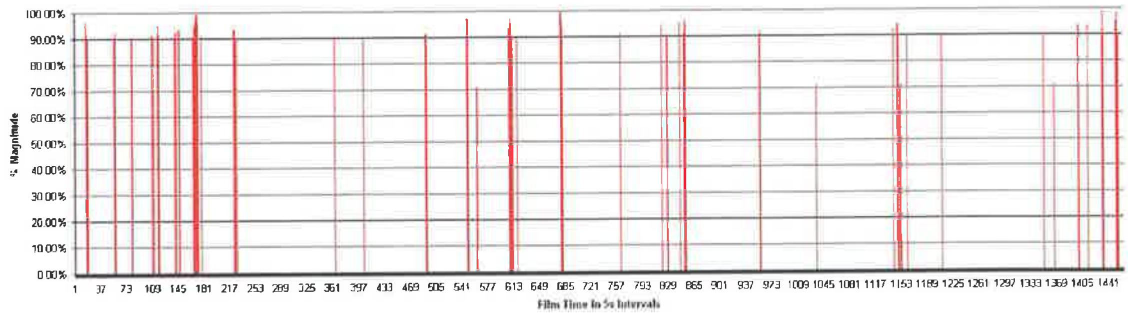
Harry Potter III P30:- HR Peaks



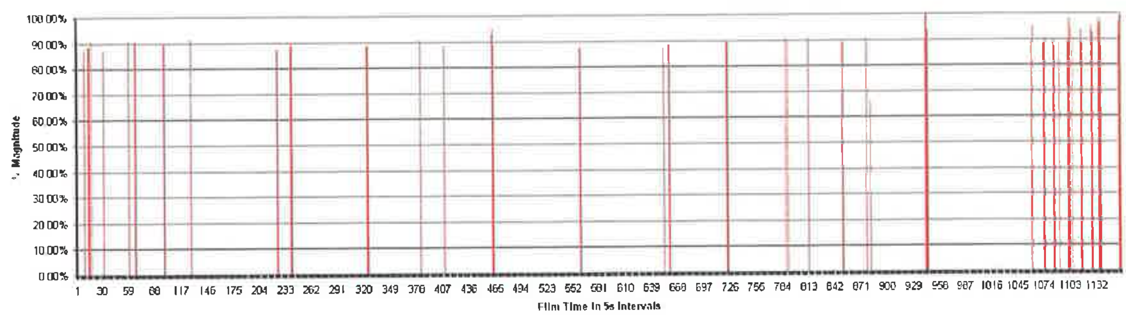
Indiana Jones III P09:- HR Peaks

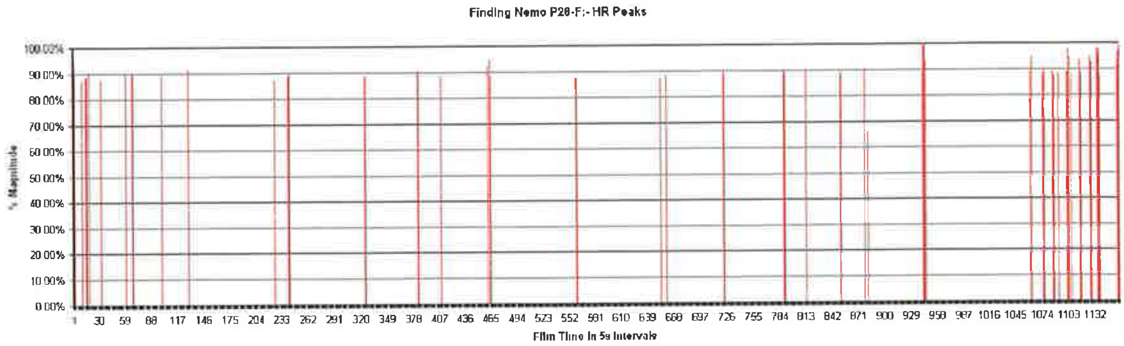


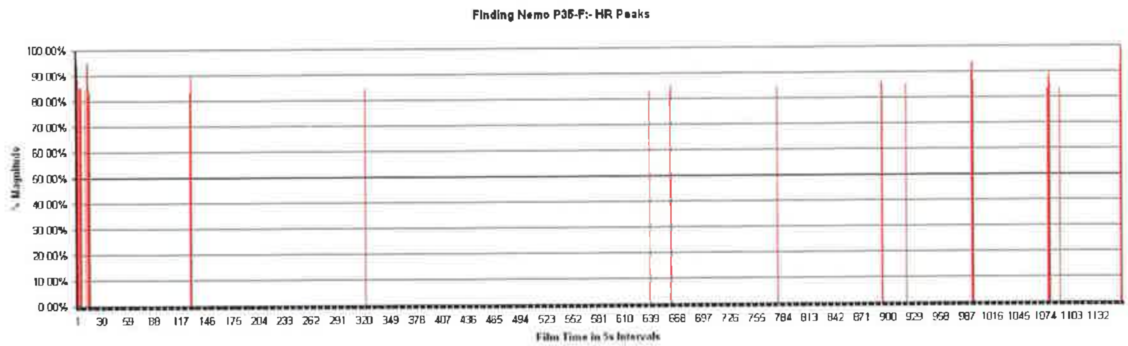
Indiana Jones III P37:- HR Peaks

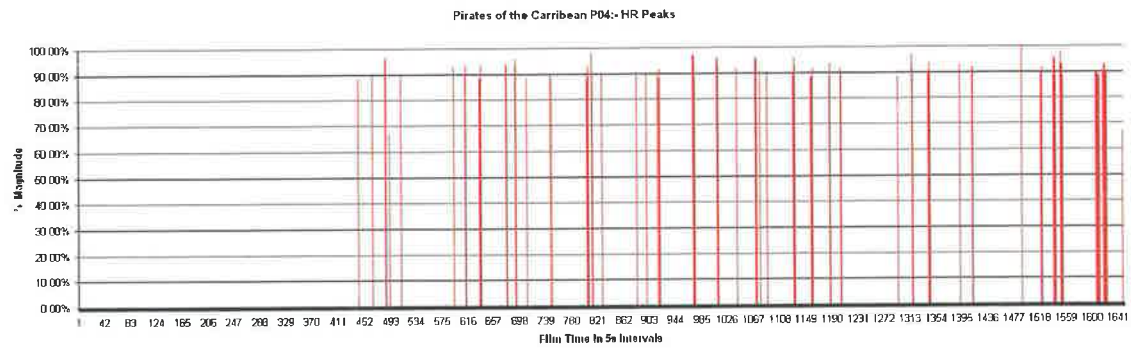


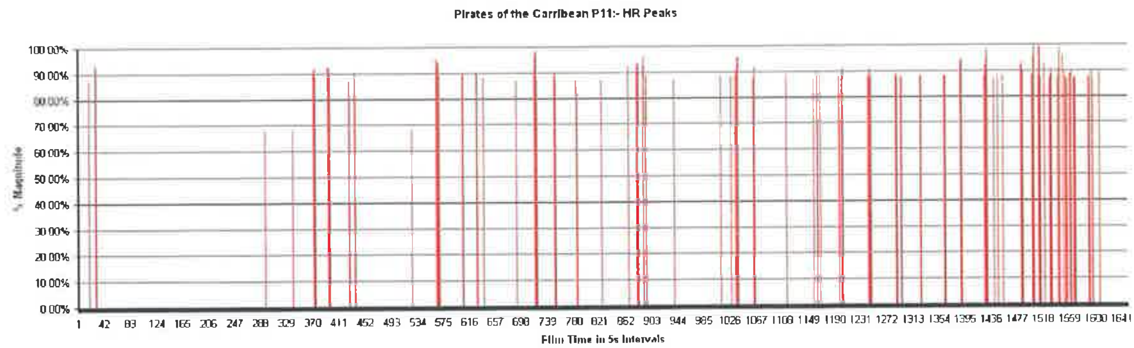
Finding Nemo P10-F:- HR Peaks




Finding Nemo P28-F:- HR Peaks
% Magnitude
Film Time in 5s Intervals


Finding Nemo P35-F:- HR Peaks
% Magnitude
Film Time in 5s Intervals


Pirates of the Carribean P04:- HR Peaks
% Magnitude
Film Time in 5s Intervals


Pirates of the Carribean P11:- HR Peaks
% Magnitude
Film Time in 5s Intervals

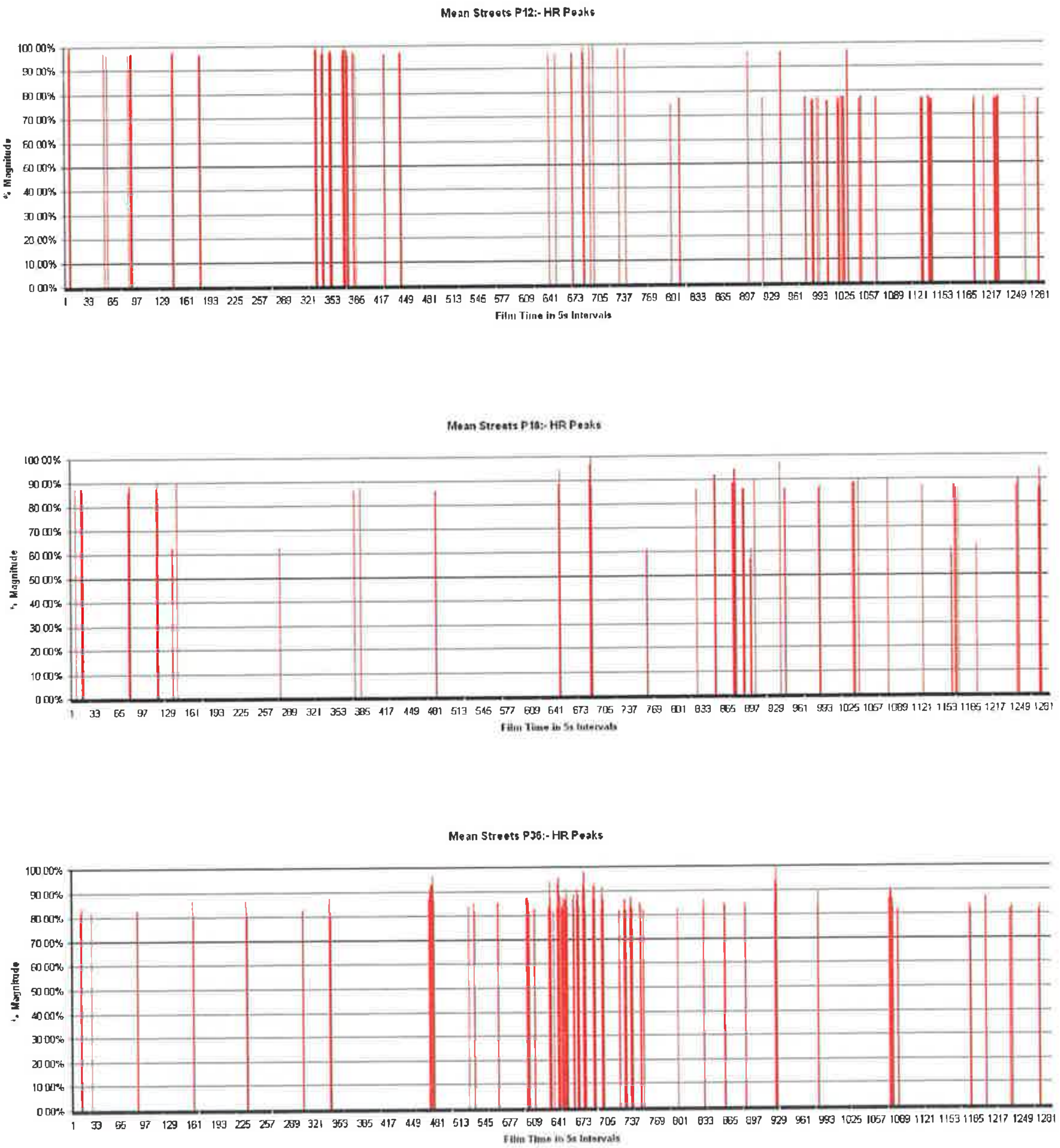


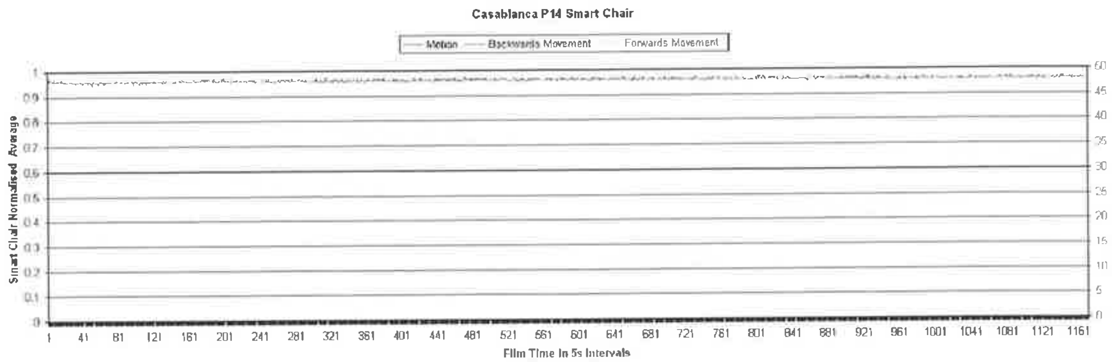
Figure 3.3(d): Heart-Rate Peaks Relative Magnitude per Viewing

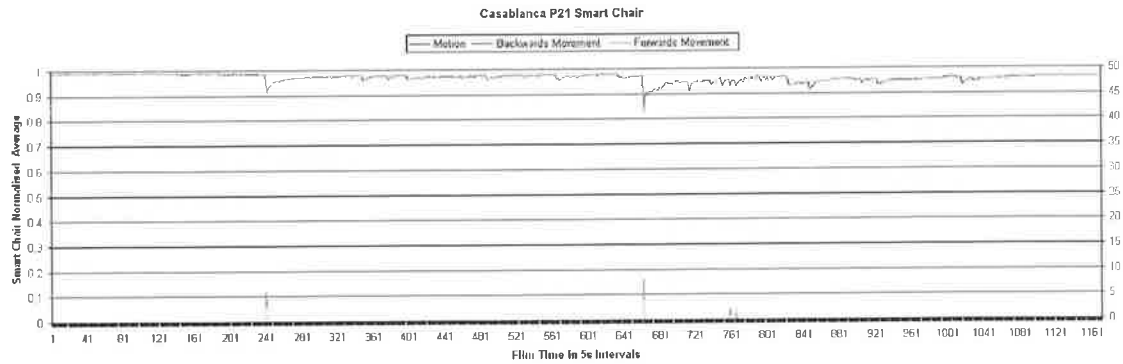
3.3.3 Motion Peak Detection

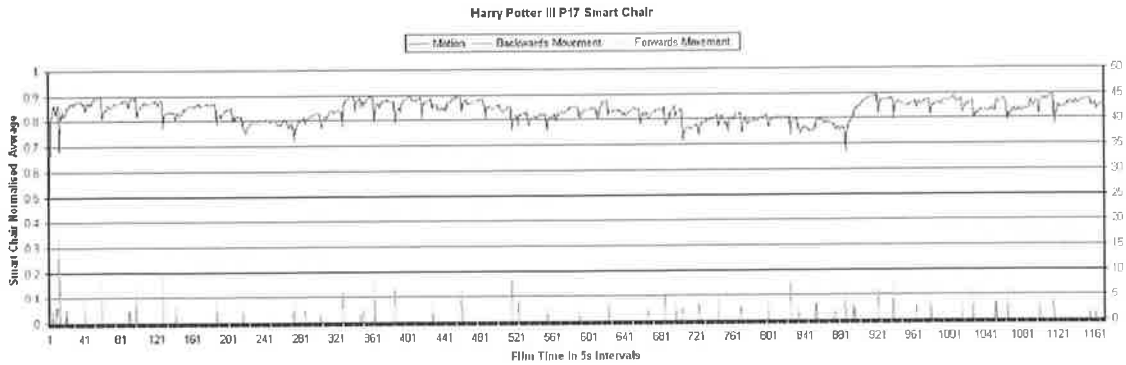
In order to detect peaks in the motion data for each viewing we first calculated the mean and then normalised the values to one (see Chapter 6.2.2). Using a sliding window approach, for each value we calculated the difference between it and its predecessor, and it and its successor. If the successive difference was significantly larger than the antecedent difference then a significant movement was deemed to have occurred. We could also identify whether it was a movement forward, away from the chair, or backwards, into the chair. The number of peaks that were present in each set of motion data are detailed below in Table 3.3(c) and the results are illustrated below in Figure 3.3(e). In four cases, P14 in *Casablanca*, P30 in *Harry Potter III*, P09 in *Indiana Jones III*, P36 in *Mean Streets*, no discernible movement was detected. This total lack of variation in these readings may be due to a malfunction in the Smart Chair. Once the motion data had been processed for the occurrence of physiological peaks, the magnitude, or relative size, of each peak was calculated as a percentage of the maximum and the results were plotted below in Figure 3.3(f).

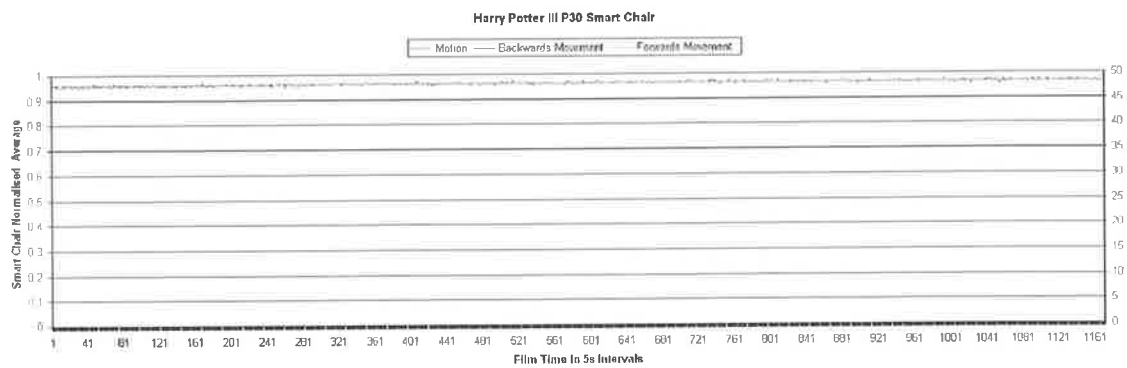
Film	Viewer	No. Peaks	No. Peaks >25% Magnitude	No. Peaks >50% Magnitude
Casablanca	P14	0	0	0
	P21	4	0	0
Finding Nemo	P35-F	n/a	n/a	n/a
	P28-F	n/a	n/a	n/a
	P23-F	n/a	n/a	n/a
	P10-F	n/a	n/a	n/a
Harry Potter III	P30	0	0	0
	P17	78	3	0
Indiana Jones III	P09	0	0	0
	P37	30	2	1
Mean Streets	P12	58	10	5
	P18	43	18	9
	P36	0	0	0
Pirates of the Caribbean	P04	11	2	0
	P10-F	82	29	6
	P11	40	12	2

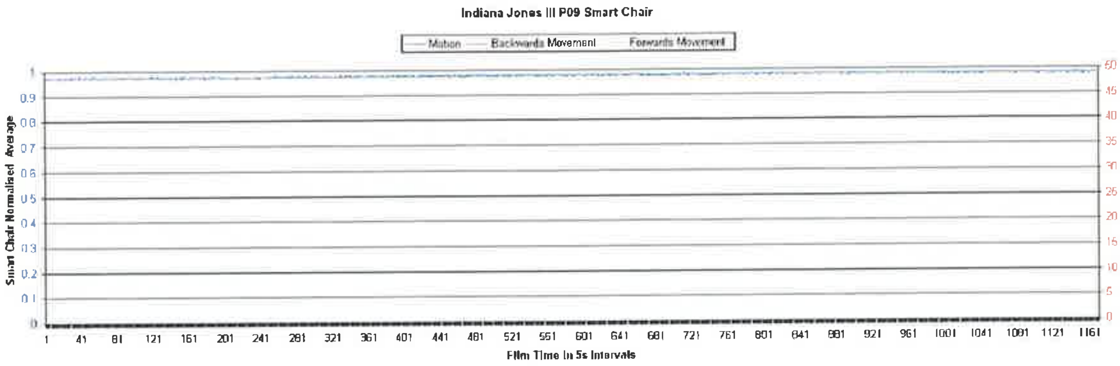
Table 3.3(c): Peak Detection in Motion Data per Viewing

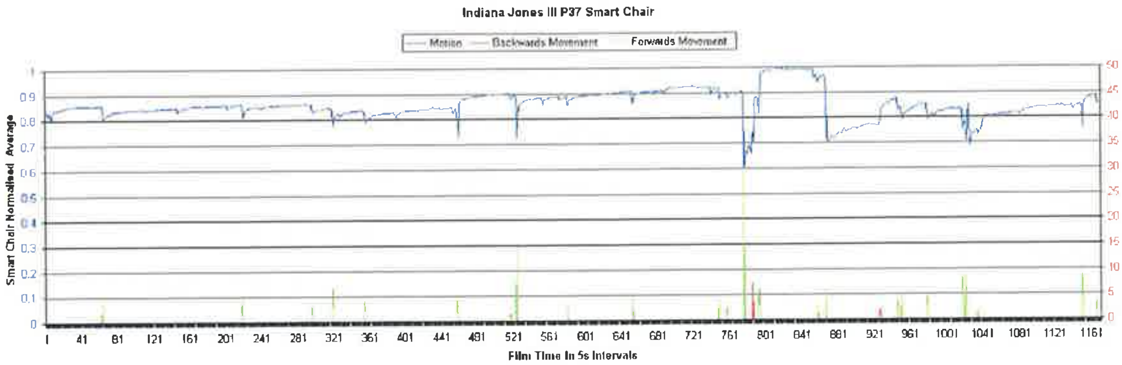

Casablanca P14 Smart Chair
Motion
Backwards Movement
Forwards Movement
Smart Chair Normalised Average
Film Time in 5s Intervals

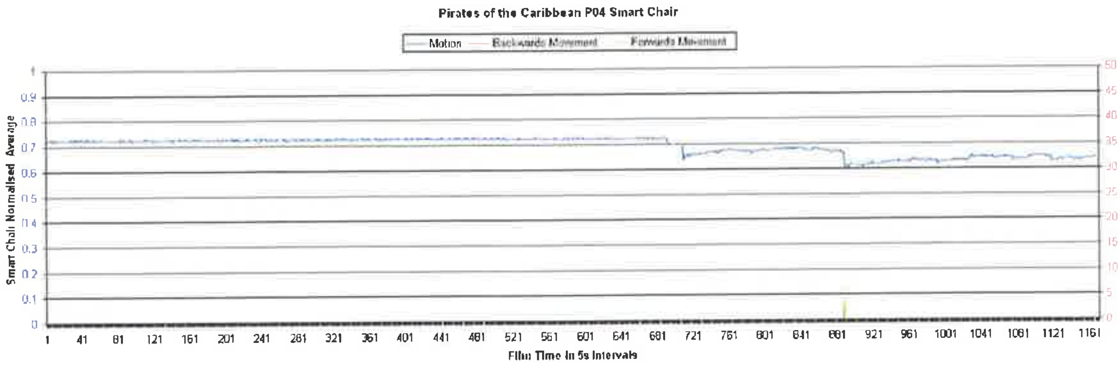

Casablanca P21 Smart Chair
Motion
Backwards Movement
Forwards Movement
Smart Chair Normalised Average
Film Time in 5s Intervals

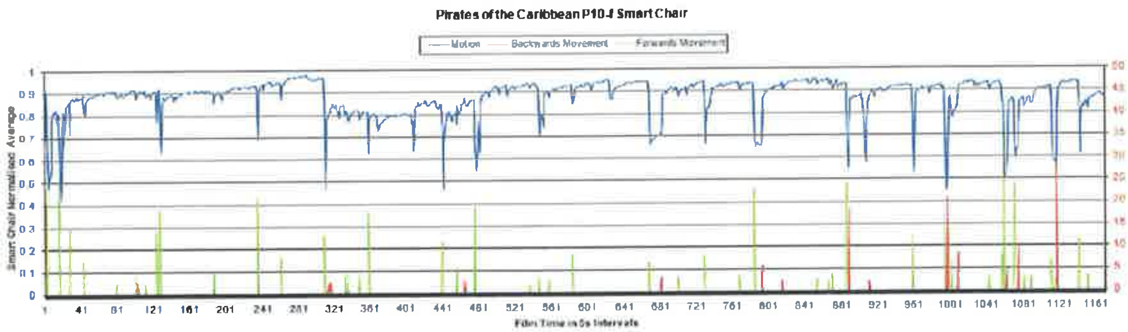

Harry Potter III P17 Smart Chair
Motion
Backwards Movement
Forwards Movement
Smart Chair Normalised Average
Film Time in 5s Intervals


Harry Potter III P30 Smart Chair
Motion
Backwards Movement
Forwards Movement
Smart Chair Normalised Average
Film Time in 5s Intervals


Indiana Jones III P09 Smart Chair
Motion
Backwards Movement
Forwards Movement
Smart Chair Normalised Average
Film Time in 5s Intervals


Indiana Jones III P37 Smart Chair
Motion
Backwards Movement
Forwards Movement
Smart Chair Normalised Average
Film Time in 5s Intervals


Pirates of the Caribbean P04 Smart Chair
Motion
Backwards Movement
Forwards Movement
Smart Chair Normalised Average
Film Time in 5s Intervals


Pirates of the Caribbean P10-I Smart Chair
Motion
Backwards Movement
Forwards Movement
Smart Chair Normalised Average
Film Time in 5s Intervals

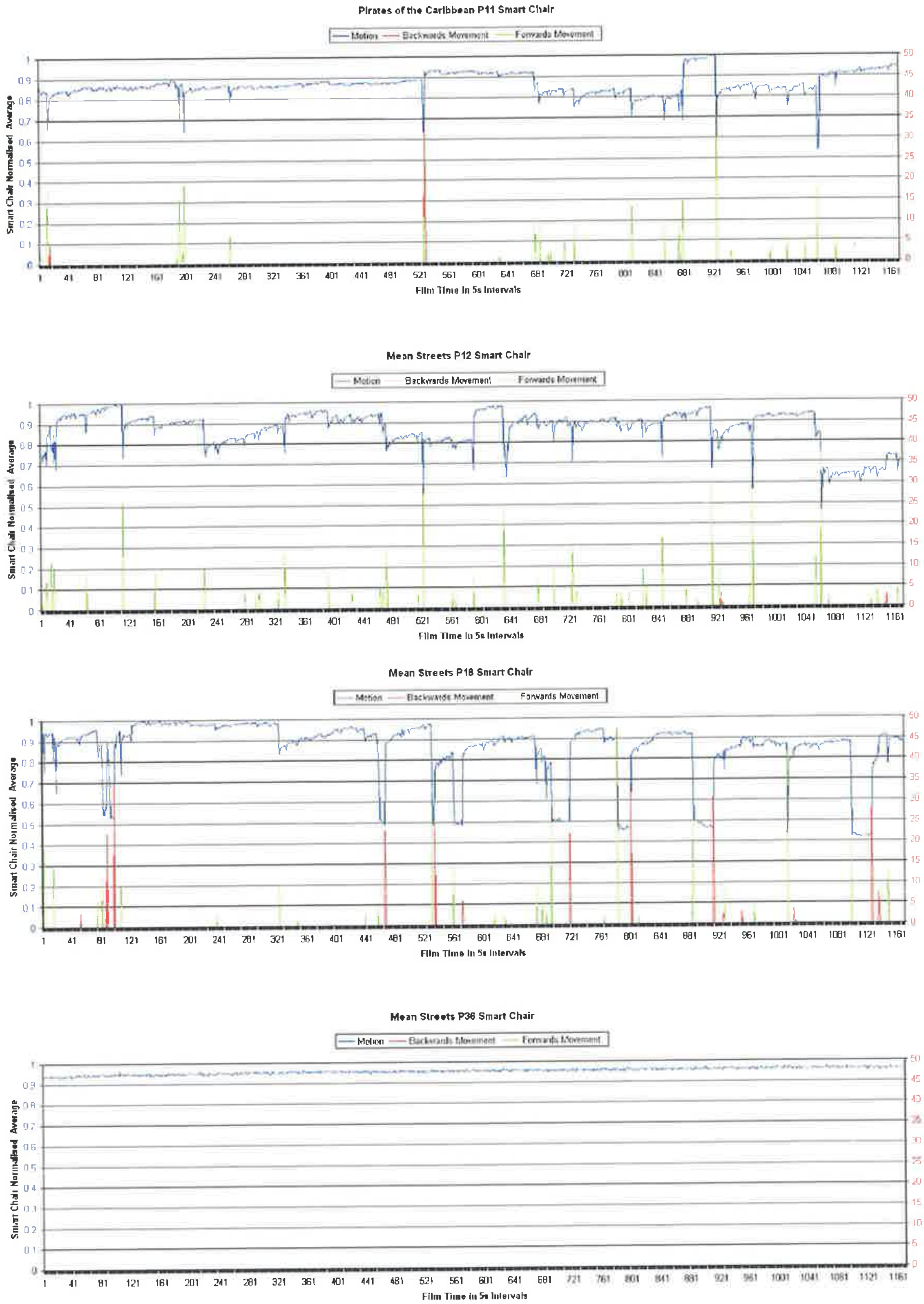
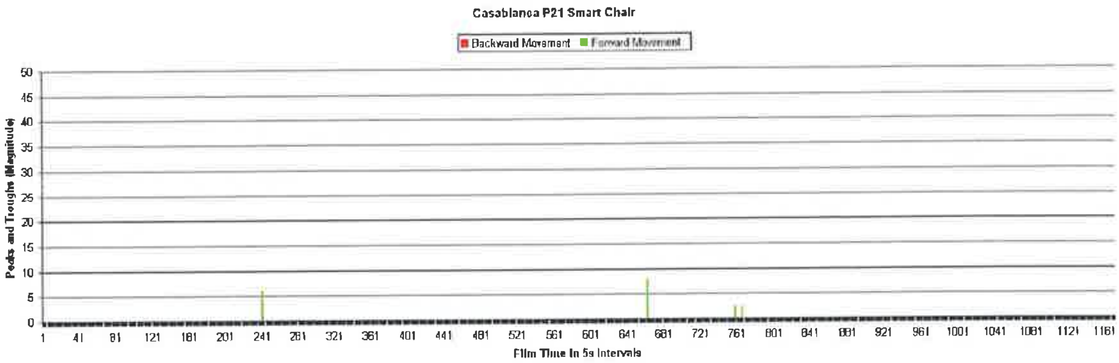
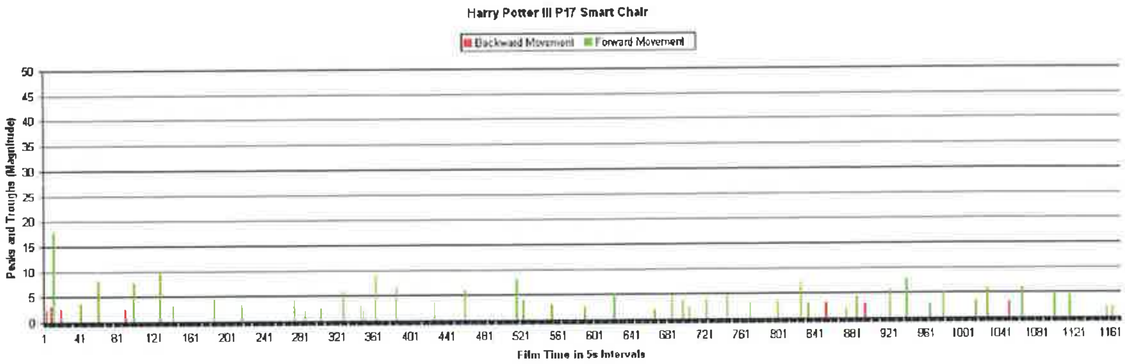
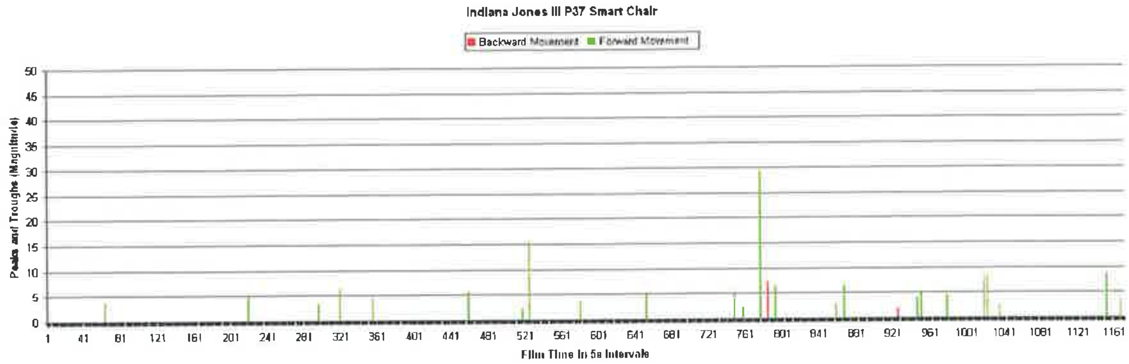
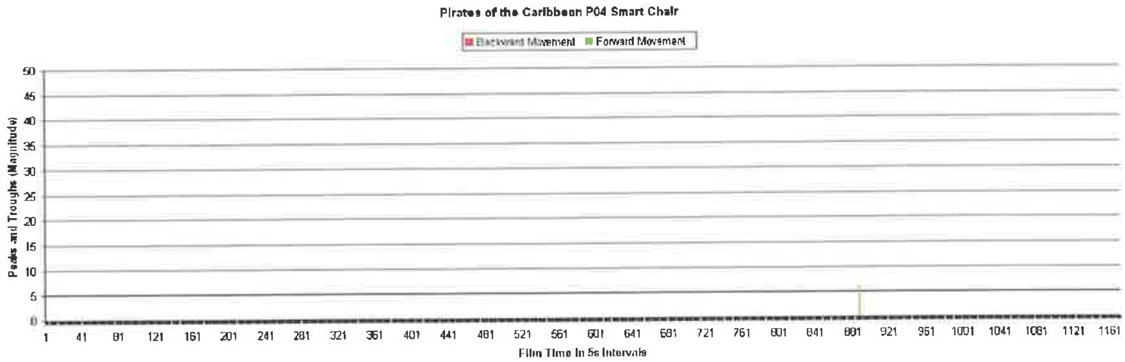


Figure 3.3(e): Motion Peak Detection per Viewing


Casablanca P21 Smart Chair
Backward Movement
Forward Movement
Peaks and Troughs (Magnitude)
Film Time in 5s Intervals


Harry Potter III P17 Smart Chair
Backward Movement
Forward Movement
Peaks and Troughs (Magnitude)
Film Time in 5s Intervals


Indiana Jones III P37 Smart Chair
Backward Movement
Forward Movement
Peaks and Troughs (Magnitude)
Film Time in 5s Intervals


Pirates of the Caribbean P04 Smart Chair
Backward Movement
Forward Movement
Peaks and Troughs (Magnitude)
Film Time in 5s Intervals

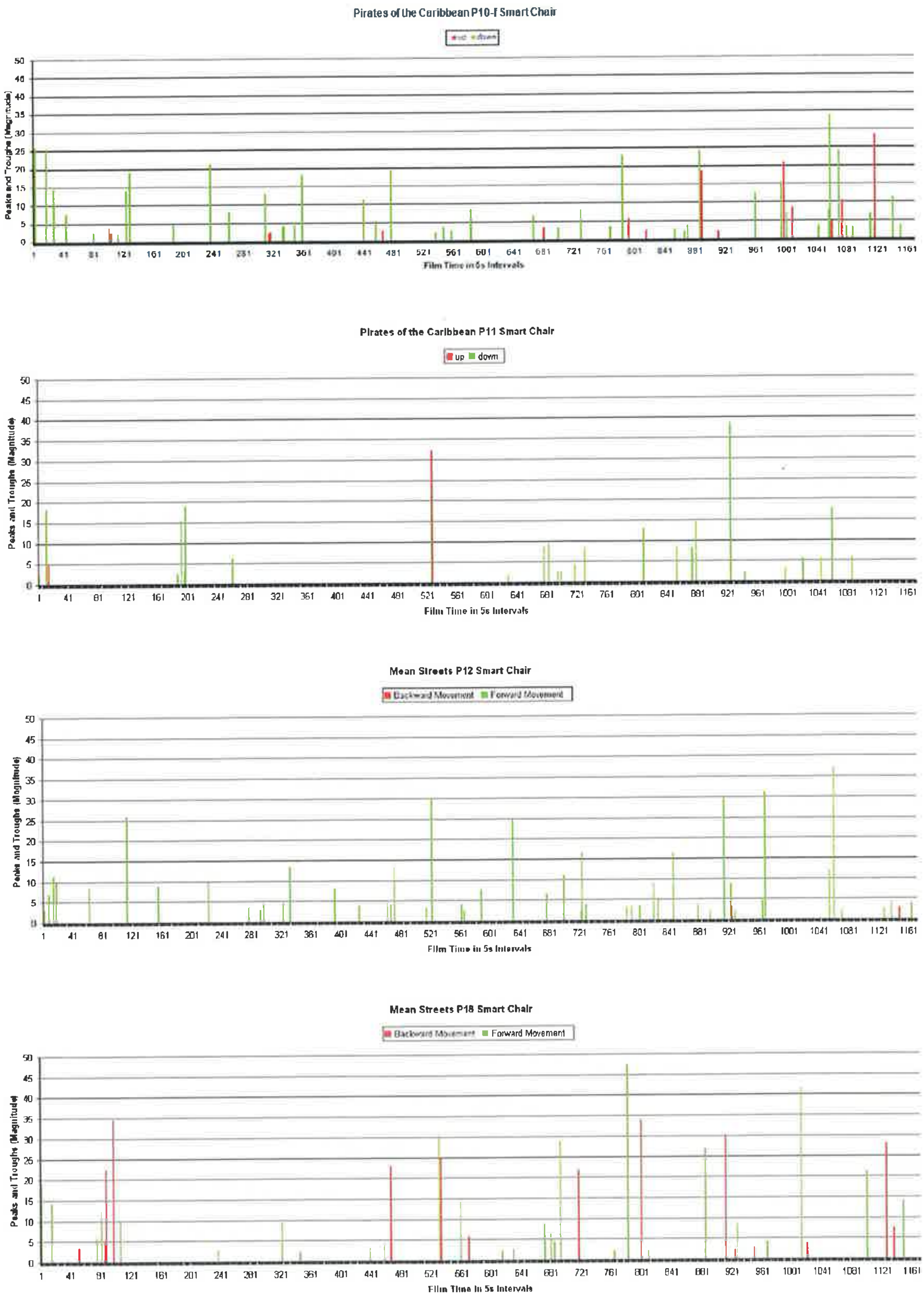


Figure 3.3(f): Motion Peaks Relative Magnitude per Relevant Viewing

3.3.4 *Summary of Biometric Peak Detection*

The results of the biometric peak detection process in terms of the number of peaks actually detected are summarised below in Table 3.3(d) and partially illustrated in Figure 3.3(g). We calculated the average number of peaks detected for each biometric, and while we see that GSR has the highest average of peaks detected, very few peaks were detected in the motion readings. However if we remove the zero values resulting from the previously mentioned non-responses in some four cases, the average number of peaks in motion rises from 29 to 43. We also see that all heart-rate peaks detected are of a high magnitude, indicating that there is no great difference in the actual size of the peaks in relation to each other. We also see that the vast majority of detected GSR peaks have a less than 25% relative magnitude, indicating that there a lot of small GSR peaks and a few large peaks.

Film	Viewer	No. Peaks			No. Peaks >25% Magnitude			No. Peaks >50% Magnitude		
		GSR	HR	Motion	GSR	HR	Motion	GSR	HR	Motion
Casablanca	P14	83	53	0	21	53	0	7	53	0
	P21	93	67	4	20	67	0	3	67	0
Finding Nemo	P35-F	59	33	-	11	33	-	5	33	-
	P28-F	253	61	-	0	61	-	0	61	-
	P23-F	218	-	-	6	-	-	2	-	-
	P10-F	37	69	-	5	69	-	2	69	-
Harry Potter III	P30	140	72	0	4	72	0	1	72	0
	P17	92	62	78	7	62	3	3	62	0
Indiana Jones III	P09	150	51	0	4	51	0	3	51	0
	P37	93	75	30	9	75	2	7	75	1
Mean Streets	P12	115	98	58	13	98	10	3	98	5
	P18	84	63	43	10	63	18	3	63	9
	P36	131	116	0	34	116	0	8	116	0
Pirates of the Caribbean	P04	78	91	11	9	91	2	5	91	0
	P10-F	132	-	82	16	-	29	7	-	6
	P11	137	126	40	10	126	12	5	126	2
Average Number of Peaks Detected		**118**	**74**	**29**	**11**	**74**	**6**	**4**	**74**	**2**

Table 3.3(d): Biometric Peak Detection Overview

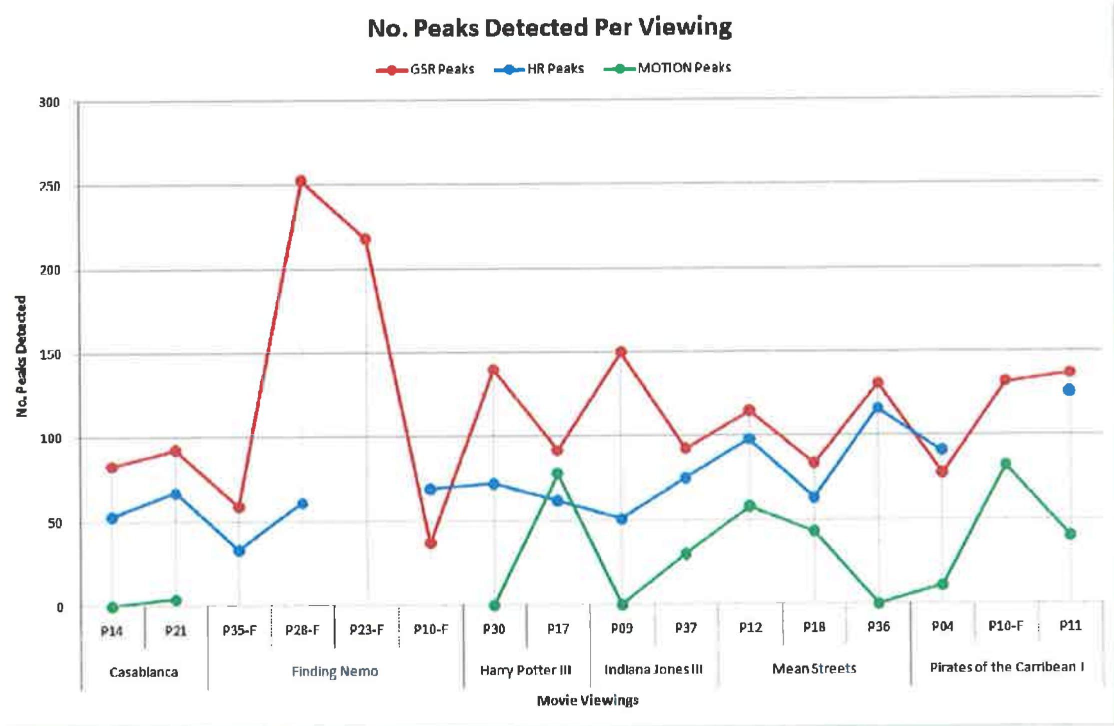


Figure 3.3(g): Overview of No. Biometric Peaks Detected Per Viewing

In this chapter we discussed the CDV*Plex* experiment conducted in order to capture, measure and detect viewers' physiological reactions to watching movies in a controlled environment. We discussed the use of sensors to monitor and record viewers' reactions. We detailed the range of biometric measurements utilised in the CDV*Plex* and the bio-sensor devices used to detect these measurements. We then discussed the details of the CDV*Plex* experiment set-up – the "cinema" laboratory, the movies chosen for viewing, details of the population of volunteer subjects and finally the experimental procedure. We then discussed and illustrated in detail, the biometric peak detection process by which "highlights" or significant changes in the biometric readings were identified as physiological reactions. This experiment provided us with a ground-truth of viewers' physiological reactions to emotional stimuli in film. In the next chapter we will discuss the audio-visual analysis of movies, with respect to movie analysis research conducted within our research group in Dublin City University.

CHAPTER 4

AUDIO-VISUAL ANALYSIS OF FEATURE FILMS

This chapter describes research conducted within the Centre for Digital Video Processing and how it pertains to the research described in this work. In developing an approach for automatically indexing the content of feature films, the common techniques used to create feature films were examined in order to derive a "grammar" for feature films which was then used to extract knowledge about what occurs in a film. This derived grammar is combined with automatic audio-visual analysis of the feature film, which is based on the extraction of low-level descriptive features, in order to facilitate the automatic detection of certain types of high-level semantic movie events.

The previous chapter described the "CDV*Plex*" experiment, which was conducted in order to compile a ground-truth of human subjects' responses to stimuli in feature films. The data resulting from this experiment shows how movie viewers react, on a physiological level, to the emotional stimuli provided by the events and devices in the feature film. Chapter 2 discussed how the device of music is used in feature films, and to what effect. Essentially, music is used to enhance the emotional impact of filmic events, and to illustrate and accentuate on-screen action; its psychological effect is considerable if largely sub-conscious.

This chapter discusses the types of events that occur in feature films and how they can be automatically detected, based on the study of film-making techniques and the use of extracted audio-visual features. Developed by Dr. Bart Lehane [Lehane, 2006] [Lehane et al., 2006] in conjunction with our research group [Centre for Digital Video Processing (CDVP), 2006], the work described in this chapter culminated in an event detection system with an interface which facilitated the event-based indexing, browsing and searching of entire movies.

For the purposes of this work however, the area of most interest is the actual detection of the movie events. The aim is to investigate any correlation that may exist between audience reaction and particular filmic events. The next section discusses film structure and explains the role of events in the context of feature films. The subsequent section then describes some filmmaking techniques and how they can enhance the quality of event detection in feature films. The event classes are themselves described, this is followed by an overview of the event detection system developed by Lehane [2006], and its use in providing indexing, browsing and search and retrieval functionality. In the final section, the audio-visual feature extraction and classification techniques used by the system's modules are described.

4.1 FILM GRAMMAR

The principal aim of Lehane's [2006] work was to develop an approach for automatically indexing movies by examining the underlying structure or film-making "grammar", and extracting knowledge about the feature film based on this structure, to allow for efficient location of segments of interest. This section discusses the definition of film events and how they can be described in terms of this film grammar.

4.1.1 Film Grammar and Events

A *frame* is the smallest unit possible in a film and as a rule movies are shot, and shown, at rate of 24 frames per second in order to ensure that there is no discernible flickering between frames when being viewed. A *shot* then is a series of sequential frames shot in one continuous run from a single camera. A *scene* is traditionally the highest-level film unit and is basically a series of connected shots. Whereas shots can be quite easily delineated by the fact that it consists of a single camera shoot, scenes are a more abstract concept and hence are somewhat ill-defined. It can be described as a film segment portraying action in one place and time, or two or more related or simultaneous actions "cross-cut" or interleaved with each other.

Film Events

A major focus of Lehane's work was creating an event-based ontology relevant to feature films. According to Lehane [2006] an event is "something which progresses the story onward" and defines it as "a sub-division of a scene that contains something of interest to a viewer" and with a minimum length of five shots. An event could be a conversation between a group of characters, or a shot sequence of people dancing. Events are the components which make up a scene, where a scene can contain a number of different event types, but a single event has a maximum length of one scene. For example, a single scene could begin with a *dialogue event* involving ten shots of people talking, followed by an *exciting event* as a fight breaks out among these people for the next twelve shots, with the scene ending with another dialogue event as the people talk again for the final eight shots (these event types are discussed in Section 4.2.2 below). Figure 4.1(a) below illustrates Lehane's conception of feature film structure.

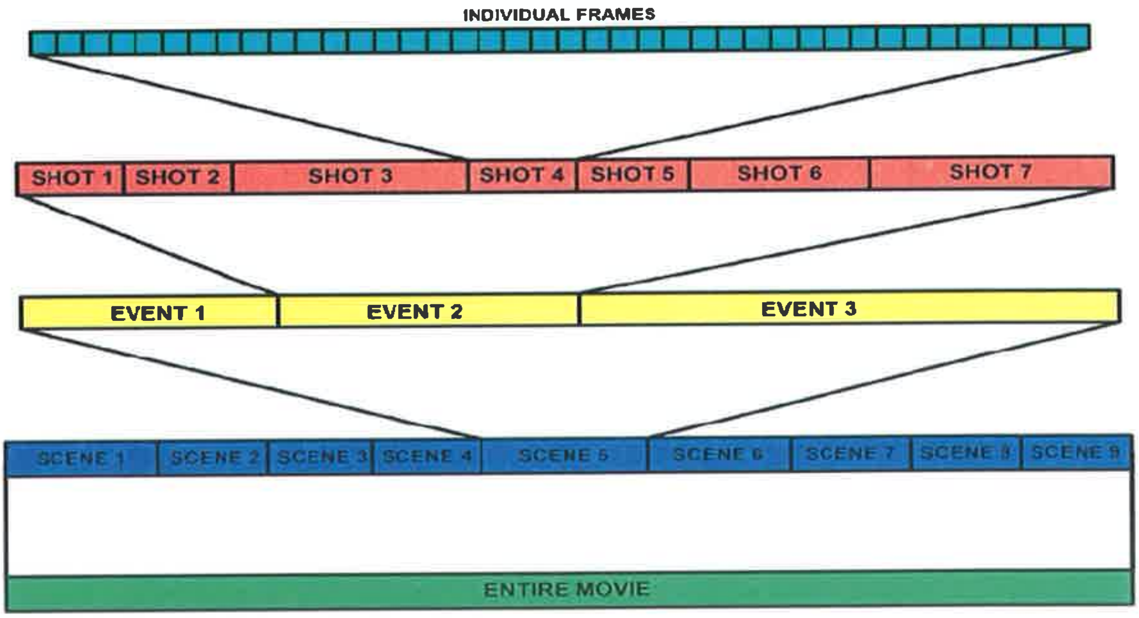


Figure 4.1(a): Lehane's Structure of a Feature Film [2006]

Events are the pieces of a feature film that, when recalling a movie they have watched, a viewer will remember as a meaningful unit; that is, a semantically coherent portion of the film. For example, a car chase is remembered as a "car chase" rather than 50 individual shots of moving cars. When viewed out of context, it would be difficult or even impossible to deduce from a single shot whether or not a car chase was taking place. It is only when it is viewed as a part of the whole event that its meaning becomes clear.

4.1.2 Scene-based vs. Event-based Structures

Conventionally, when imposing a higher-level structure on video whether it is scene-based or, in this case, event-based, shot boundary detection techniques are the most commonly used first step. An extensive review of past and present approaches to shot and scene boundary detection is given in [Lehane, 2006, Section 1.4]. As can be seen from this, a more descriptive *video-scene-shot* representation is preferable to a simpler *video-shot* representation, as detecting scene boundaries does improve navigation and enhance knowledge of film structure. However, even by using this scene-based structure no knowledge is obtained regarding the content of the scene. This is a major disadvantage to a user who is searching for specific clips of the film, and in browsing terms, gives no advantage over the unwieldy shot-based systems. Lehane surmised that event-based structures would provide enough information to facilitate this functionality but that existing event-detection techniques, while they utilised conventional film grammar rules, were hindered by a lack of proper understanding for the process of film creation. Also he discovered that little research has been conducted in detecting all relevant events across entire feature films, with even less work having been undertaken in search and retrieval of specific events from within a film based on the user's requirements.

4.2 EVENT DETECTION

Filmmaking is a highly creative process and the underlying concepts initially devised by early filmmakers have become today's ubiquitous filmmaking building-blocks. These are the basic ingredients of the undoubtedly innovative and highly creative film-making process which produces interesting feature films. Viewers have now come to expect the use of these techniques and experience confusion or become disoriented if these conventions are not followed. The event-based ontology formulated by Lehane is based on analysis of these film creation techniques.

4.2.1 Filmmaking Techniques

Lehane [2006] hypothesised that, in order to properly analyse the movie content, it was essential to have an appropriate understanding of these underlying creative techniques as used by film-makers. To that end, the methods commonly used by directors, editors and sound engineers were investigated [Lehane, 2006, Chapter 2] and were a major factor in the design of the event detection techniques. Thus knowing the well-defined directing and editing principles employed to influence audience reaction, and detecting when these devices are used, it is possible to infer knowledge about the event, and indeed the film.

Camera placement can indicate which shots may be part of the same event. *Shot repetition* can indicate interaction between multiple characters. *Lighting and colour changes* can indicate a new event has begun, as they will typically consistent throughout an event. Several factors can be used to differentiate between relaxed and exciting happenings. Low levels of *camera movement* suggest relaxed activities on-screen, whereas high levels indicate excitement. Similarly for *object motion* or the movement of objects and characters as captured within the screen. Faster *editing pace* and therefore shorter shot length implies intensity, with the opposite being true for slower pacing. Finally, the *audio type and composition* present can assist in event detection as the presence of speech indicates a character is speaking, while the presence of music or silence can imply a musical and/or emotional event is occurring.

Every film-maker will shoot an event differently and therefore, rather than try to detect the event as a whole, it is more effective to detect the set of underlying components which constitute the event, at least some of which must be present for positive classification. The filmmaking devices, such as those mentioned above, which make-up these components are detected by automatic audio-visual analysis tools and, based on this extracted knowledge, the film is indexed by imposing an event-based structure on it. Audio-visual analysis allows for the maximum amount of information about the content of the movie to be extracted, provided the appropriate set of features are selected and implemented. A set of data classification algorithms then uses the extracted feature data to detect event components and classify event occurrences accordingly.

4.2.2 *Event Classes*

An almost infinite range of possible events can occur in any movie and attempting to create a set of detectors for each one of these is not only impractical but unrealistic. Consequently, the event types, or classes, should be restricted in number and broad enough in definition to allow for straightforward browsing, and yet sufficiently describe the relevant content. Lehane defined three event classes in his ontology, namely: *exciting events*, *dialogue events*, and a superset of different event classes known as *montage events*.

Dialogue Events

Dialogue forms a major part of any movie, and is often the primary vehicle for conveying information to the viewer on the plot and background of the film. Dialogue events include people talking, such as conversations between two or more characters, a person addressing a crowd or a teacher addressing a class. Dialogue events tend to have high amounts of clearly audible speech. When shooting a conversation between characters a director will often use a static camera, fixed on the people talking, thus dialogue events will usually a lot of shot repetition as the camera focuses in turn on each individual speaking [Lehane et al., 2004a, 2005]. Therefore, the audio-visual features most useful in detecting dialogue events are: a measure of the camera movement, a measure of the amount of speech and a measure of the amount of shot repetition.

Exciting Events

Also referred to as *action events*, these typically occur less often than dialogue events but can form central components of many films. Exciting events include fights, car chases, battles etc. More open to interpretation than the clearly defined dialogue class, an exciting event is defined as "a sequence of shots that makes the viewer excited" [Lehane, 2006, p.42]. Exciting events tend to contain a lot of movement and as the camera moves from one location to another a low amount of shot repetition is also usual. Increased editing pace resulting in decreased shot lengths is also typical [Lehane et al., 2004b]. Therefore, the audio-visual features most useful in detecting exciting events are: a measure of motion in the video, a measure of the amount of shot repetition, and a measure of the editing pace.

Montage Events

This event class is a superset of a number of different event types, the common denominator being the existence of a dominant musical presence on the accompanying soundtrack, or at least the absence of speech. First of all is the actual *montage* event where spatially and/or temporally discontinuous shots are interleaved, typically with a constant musical accompaniment, causing viewers to alter their perception of the seemingly unrelated actions in these shots not as discrete occurrences but as part of a semantic whole. In addition to this, the montage event class definition has been expanded to include an *emotional* event type, examples of this being a shot sequence of somebody crying or a romantic embrace or occurrence. Emotional events and montages are closely linked as many montages have strong emotional associations and undertones. The final addition to the montage superset is the *musical* event type, examples of which are a band playing a live song on-screen or a musician playing at a funeral. Typically, occurrences of musical events are relatively rare.

These montage type events contain non-speech audio which is usually music, and occasionally silence, with low amounts of camera motion, and relatively slow paced editing. Therefore, the audio-visual features most useful in detecting montage events are: a measure of the amount of non-speech audio, a measure of the amount of camera movement, and a measure of the editing pace.

4.2.3 Uses for Automatic Event Detection

A system was built to utilise this new event-based approach and detection process, incorporating a number of useful areas of functionality, which are detailed below. An interface was devised to incorporate this functionality by presenting the detected events for browsing, and also allowed searching and retrieval of the film events.

Automatic Indexing

Indexing is achieved by analysing the underlying structure of the film and extracting content-based knowledge. For the purposes of automatically indexing film content where an event is considered the fundamental unit of retrieval, all of the applicable events in the film should be detected and classified according to the devised ontology [Lehane &

O'Connor, 2006]. It is important to allow for different user interpretations of film content. For example, where an exciting argument is taking place, the segment is classified as both a dialogue event and an exciting event so that no matter what the human interpretation, a user can find their event.

Event-Based Browsing

One of the main aims of Lehane's work was to build a system that could present the indexed version of a film to a user in a useful and accessible manner, in order to allow them to browse an event-based index of a film, and view and locate specific events within it.

Event-Based Searching

Another primary goal of this research was to build a system which facilitated the searching of movie content which would allow for a type of user-defined browsing through a film. This would allow for a "tailored" retrieval of events using audio-visual information extracted as descriptive features, the objective being to allow a user to select features most likely to feature in the event they sought.

Event-Detection System

The actual event-detection system designed by Lehane is illustrated in Figure 4.2(a). The system consists of three modules. Low-level feature extraction extracts the features necessary for event component identification. Shot-level feature vector generation aggregates these low-level features at the common data unit of a shot. This allows shots to be directly compared to each other. High-level event detection creates the event-based index for the film by automatically classifying the generated features. The techniques used by these modules will be further explained in the following section.

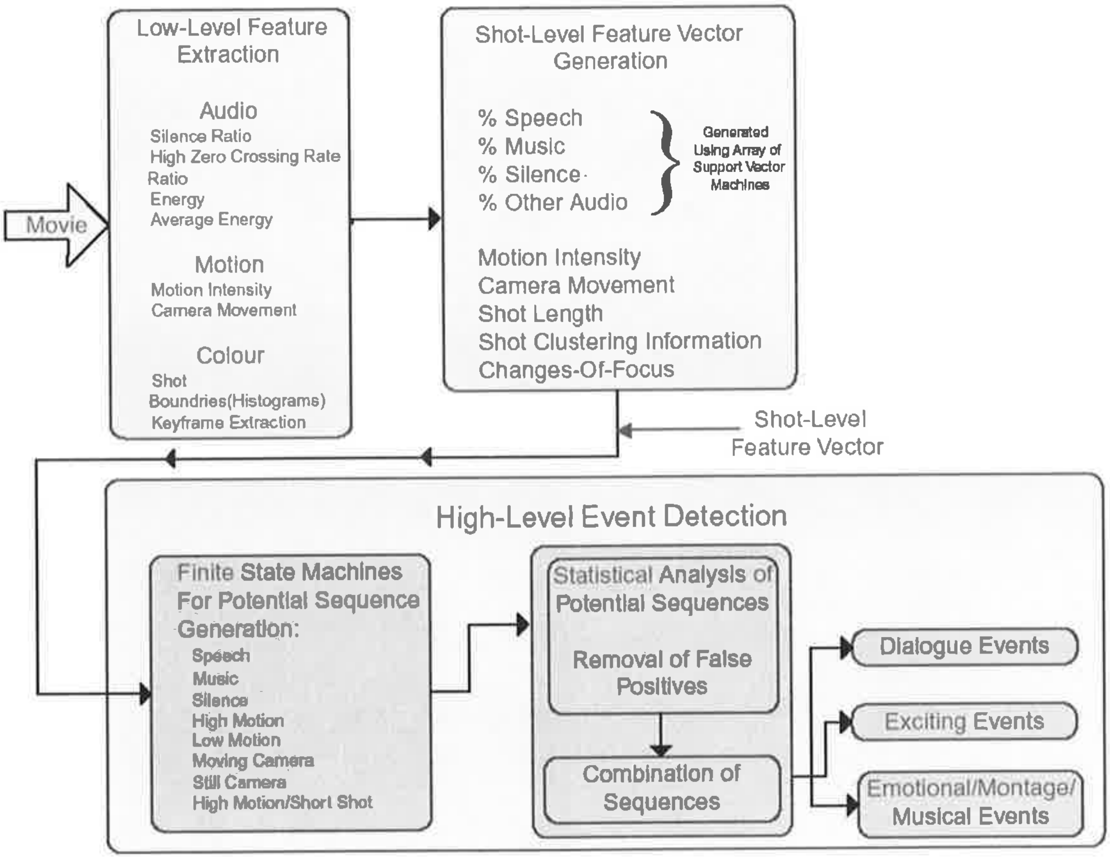


Figure 4.2(a): Lehane's Event Detection System [2006]

4.3 AUDIO–VISUAL ANALYSIS

As previously stated the aim of the work described in this chapter is to automatically detect events in a feature film, which is achieved using audiovisual analysis based on filmmaking techniques as previously described. This audiovisual analysis involves detecting appropriate features, and then applying data classification methods to extract knowledge about the film. As described in Section 4.2, Lehane has devised three event classes namely *dialogue events*, *exciting events* and *montage events*. Detection of these events is based on the fact that a number of components occur frequently within the events of each class, although there is some variation between the events within each class. For example, not all dialogue events are shot in the same way.

4.3.1 Low-Level Features

As mentioned above in Section 4.2.2, the features required to detect the three event classes based on filmmaking components are: a description of the audio content (where the audio is placed into a specific class; speech, music etc.), a measure of the amount of camera movement, a measure of amount of motion in the frame (regardless of camera movement), a measure of the editing pace, and a change-of-focus detector. The latter feature is important as it allows for the detection of the boundaries between events and therefore where the film changes focus from one event to another. This could be between two event classes (for example, where a dialogue event finishes and a montage event begins), or within the same event class (for example, where one dialogue event finishes and another dialogue event begins). These low-level features can be grouped into three categories, namely colour-based, motion-based and audio-based.

Colour-based Low-Level Features

Colour-based analysis is used to extract a feature from a video frame so that it can be compared to other frames to facilitate frame matching and contribute to shot boundary detection. There are several methods for comparing frame images. One method is to directly compare pixel values but this is quite computationally expensive. Another technique is edge detection, which involves filtering the images in order to detect edges. This is also often used in shot boundary detection, but generally colour histograms are more useful. Colour histograms are a representation of colour values throughout the entire image rather than localised pixel values. It is an accurate and efficient method for comparison of image similarity and reliable shot boundary detection

Motion-based Low-Level Features

MPEG-1 encoded video was used in Lehane's [2006] analysis work, therefore the motion vectors created during the encoding process can be used as they are acceptably efficient in capturing the movement in the video and, as they are compressed within the MPEG-1 data stream, also allow for fast analysis. An accurate measure of the motion intensity in a particular frame is the standard deviation of the motion vectors. Different motion intensity values indicate the amount of motion within a frame, which is a measure of how much the objects in the video are moving. Detecting camera movement in a particular frame is achieved by measuring the zero motion vector runs, a novel approach developed by Lehane. This approach was developed rather than use existing but computationally intensive techniques. Also this method concentrates on efficiently and accurately detecting any camera movement present, and does not extract unnecessary knowledge on camera direction. Combining these two features allows for an accurate representation of the movement present in a video.

Audio-based Low-Level Features

Four audio features were extracted and combined to generate a single representative value for each second of the audio signal associated with the video under analysis. Lehane [2006] implemented a reliable and efficient set of features, variations of which have been used successfully in previous work on audio classification. The high zero crossing rate is very useful in speech detection as speech usually contains lots of short silences in between the spoken words. The silence ratio is a measure of how much silence versus non-silence there is in an audio sample, and is useful for distinguishing between speech and music, as music tends to have a low silence ratio while the gaps between spoken words in speech means it tends to have a high silence ratio. Two energy-based features are calculated, the first being short-term energy which represents the signal's amplitude variation over time. The second is the short-term energy variation which is the ratio of low to high energy values. These features can be used to distinguish between silence and speech or music, as silence will have low energy values. Together these features are used to classify each second of the audio data into the different states of speech, music, quiet music and silence. Even within a particular state there is large variation in an audio signal and a combination of features is used to extract as much information as is possible before a classification is made.

4.3.2 Generating a Shot-Level Feature Vector

In order to discover meaningful information about the video content, a higher-level approach must be taken, where films are analysed at shot and scene level rather than at the various sampling frequencies, for example, the frame-level analysis of colour features or the one second frequency of audio analysis. Each of the low-level features described above are therefore used as the basis for event-level analysis. The next step in achieving this is the generation of a feature vector for each shot in the video [Lehane, 2006], where the vector contains a set of aggregated features indicating the percentages of speech, music, silence, quiet music, other audio, static camera frames, non-static camera frames, motion intensity and shot length.

Speech: The amount of speech present in a shot is used to determine whether characters are speaking, which is very important for detecting dialogue events

Music & Quiet Music: The amount of music, quiet or otherwise, present in a shot is used primarily to determine areas in a film soundtrack which are music-rich, which is important in detecting montage events

Silence: The amount of silence present in a shot can be used to locate areas where there is little activity in the film soundtrack.

Other Audio: The amount of non-speech/music audio present in a shot is used to identify areas where there is undetermined audio on the soundtrack.

Static Camera: The amount of static camera frames in a shot is used to locate areas where the film director uses a still camera, often used in order to allow the viewer to fully concentrate on some on-screen activity.

Non-Static Camera frames: The amount of non-static camera frames in a shot is used to locate where the film director uses a moving camera, often used to create excitement, and is therefore important in detecting exciting events.

Motion Intensity: The amount of movement within the frames of a shot. Shots with high amounts of motion intensity can be used by filmmakers to induce excitement.

Shot Length: The length of each shot is calculated from shot boundary detection. A number of short shots in succession may indicate that an editor is trying to create excitement, while long shots indicate that the editor is trying to relax the viewer.

Shot clustering involves grouping similar shots together, where the aim is to cluster shots that are filmed using the same camera in the same location. Detecting when one cluster ends and another begins makes it possible to detect changes in the focus of on-screen activity. This is very important for event detection as it is used to detect event boundaries. Also the clustering information can be used to measure the amount of shot repetition present in an event.

4.3.3 High-Level Event Detection

Detection of events in feature films is achieved in two further steps where first a set of *potential sequences*, or sequences of shots with a common characteristic, are generated and then analysed to produce a list of dialogue, exciting and montage events, by automatically removing any sequences which are false positives and combining the remainder in order. The potential sequences are located using an array of finite state machines (FSMs) [Lehane, 2006], which use the features from the shot-level feature vector described previously. FSMs are models of behaviour composed of states, transitions and actions which are used to classify data.

A potential sequence is a sequence of successive shots in which a specified feature occurs frequently. The FSMs are used to find sets of shots where particular features are dominant and tag them as potential sequences. Many potential sequences will overlap, with the result that many shots will be present in more than one potential sequence. This allows for different user interpretations of on-screen events, as previously discussed in Section 4.2, to be accommodated. But these potential sequences cannot be classed as events before some additional processing to examine how many of the shots in the sequence contain a different feature.

As aforementioned, Lehane proposes that event detection is more successful if it is carried out by detecting the conventional filmmaking components that define an event type, at least some of which are present in any given instance of that event type. For example, typically most, if not all, shots in a dialogue event will contain speech and most of the shots in an exciting event will contain high amounts of motion. But movie segments with a voice-over contain speech, but cannot be classified as conversations.

The final feature used to detect events is extracted only at potential sequence-level. The *cluster to shot ratio* (CS ratio) is the ratio of clusters in a sequence of shots, to the amount of shots in the sequence. A low number of clusters indicates a high amount of shot repetition, as many similar shots are clustered together, whereas a high number of clusters indicated that shots are not repeating too frequently. Using this information the potential sequences are combined and filtered to create the event list.

One of the aims of the work described in this thesis is to use the data gathered in the biometric experiment described in Chapter 3, and correlate it with this set of events to discover how people respond to different types of events. The results of which will be discussed in detail in Chapter 6. In the next chapter we shall describe how we annotated a number of the feature films used in the biometric cinema with emotion labels, or tags, which we will correlate with the results of the audio analysis and the results of the event detection as described in Lehane's work [2006].

CHAPTER 5

MANUAL ANNOTATION OF FEATURE FILMS

This chapter describes the process by which a selection of feature films were manually annotated for the presence of emotional content. We will first look at the set of tags used to mark up each feature film, explaining their provenance and meaning, and then progress to describing the process of annotating each film and how that process was executed. Following this, we will discuss the movie selection itself, including a synopsis of each storyline. Finally, we will examine the results of the annotation.

The previous chapter discussed work conducted within our research group involving audio-visual analysis in feature films for automatic event detection. This work has resulted in entire movies being indexed for *dialogue events* such as conversations, *exciting events* such as car chases and fights and *montages*. In order to do this a set of audio-visual features were implemented to capture information on conventional filmmaking techniques which form the basic components of movie events. By detecting the presence of these techniques and measuring the manner in which they were used and combined, it was possible to classify portions of the film as containing an event or events of a certain type. Chapter 2 described the use of music in film as a key filmmaking device, and how it can be used to influence emotional states of viewers, manipulate their perception of movie events and to generally enhance the movie viewing experience. Chapter 3 then described an experiment in which physiological reactions experienced by viewers as they watched movies were recorded by means of biometric instruments. The aim of the research described in this thesis is to investigate the correlations between viewers' reactions while watching a feature film, the types of events occurring in the film and the emotional connotations of what is occurring in the film, with particular attention to portions of the movie which are music-rich. In order to accurately identify the emotions conveyed during the film to facilitate the desired correlation analysis, a selection of movies were manually annotated for emotional content. This process is described in this chapter, including details of the set of emotion labels used to annotate the films. A description of each film selected to undergo manual annotation is also given, followed by a discussion of the results.

5.1 EMOTION TAG SET

In order to annotate the emotional content of the feature film extracts, a set of labels or tags describing emotions first had to be chosen. The following section details the emotion tag sets that were considered for this purpose, followed by the details of the set that was eventually employed.

5.1.1 Sets of Emotion Types

Russell-Mehrabian Emotions

The Russell-Mehrabian [1977] set of 151 emotion types are a detailed and descriptive set of emotion labels, however they were deemed unsuitable for our purposes as they were not concise enough, or rather they were *too* precise. It was impractical to consider using such a large number of possible classes of emotion for manual annotation, as it is too difficult to make such fine-grained distinctions quickly and efficiently.

Salway's Ortony-Based Emotion Tokens

Salway and the Information Extraction and Multimedia Group in the University of Surrey in the United Kingdom are involved in processing and generating linguistic descriptions of still and moving images, including time series, and in multimedia information retrieval and extraction. Specifically, Salway is concerned with developing intelligent multimedia information systems, such as digital image and video libraries, particularly by providing knowledge-rich representations of media content.

In recent years, Salway has conducted work on analysing feature films using the associated "Audio Description" [Salway & Graham, 2003], a facility which is now being made available for an increasing number of television broadcasts and cinema releases both in Ireland and the United Kingdom. The provision of Audio Descriptions with programmes and movies is a service developed to cater for people who are visually impaired and comprises of a spoken account describing the filmic action occurring on-screen. In effect, when a film is furnished with an Audio Description the story told by the moving image is re-told in words. The Audio Description is a script read aloud by an actor and recorded to fit around the diegetic or on-screen dialogue in the film's soundtrack, in order to minimise disruption to the film experience.

Salway's focus has been on extracting information about the emotions of the movie's characters, from this written Audio Description script, in order to understand more about the story or plot of the film. This work involved analysing feature films of different genres in order to extract emotive content information based on the associated Audio Description script text. They developed a list of indicative keywords or emotion tokens, which were grouped into 22 categories of emotion based on the emotions classified by Ortony, Clore and Coolins [1988].

Ortony gives three basic classes of emotions which are based on reactions to events, agents and objects: being *pleased* vs. *displeased* (reaction to events), *approving* vs. *disapproving* (reactions to agents), and *liking* vs. *disliking* (reactions to objects). These three basic classes are then differentiated into a number of groups of the 22 emotion types, as outlined in Figure 5.1(a).

These emotion classes and their grammatical variants were used as entry points into WordNet [2006] [Fellbaum, 1998] which was then followed by some post-processing to ensure unique mappings from emotion token to emotion type or class. That is, when detected in an Audio Description text, each keyword is indicative of just one emotion. In practice just 21 of the emotion types were used, as it was discovered that one of Ortony's emotions, *"FEARS-CONFIRMED"*, didn't map well to WordNet. In total 627 emotion tokens were selected for these 21 emotion types, and this list will be described in more detail below.

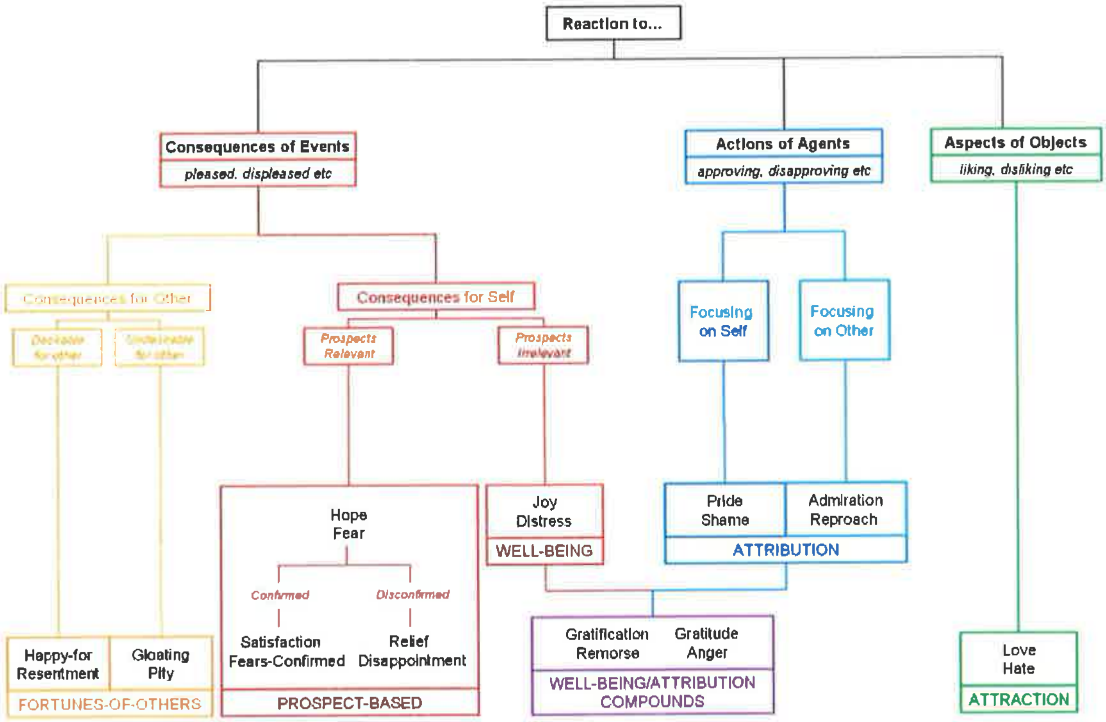


Figure 5.1(a): Ortony's Emotion Structure Framework

Salway's method for detecting the emotional content in the Audio Description, and hence visibly manifested emotional content of the film, was to scan the text of the time-coded Audio Description for these keywords and to then classify these into one of the 21 types of emotion. For example, the use of the keyword "exhilaration" in the Audio Description was purported to be indicative of a character experiencing *"JOY"*, and detection of the keyword "alarmed" is indicative of a character experiencing *"FEAR"*. The result of this analysis was the generation of a file, for each of the movies, which contained the sections of Audio Description in which an emotion had been identified, and the approximate time of when the action it describes occurred. The Audio Description refers to what is depicted on-screen at, or near to, the moment it is spoken, due to the fact that the describer must work around existing dialogue.

5.1.2 Emotion Tag Set Chosen for Annotation

Ultimately, it was decided to build the set of labels used for the purpose of manually annotating the selection of movie segments upon those used successfully in previous research carried out in feature film analysis, namely the work conducted by Salway, as described above. As well as allowing for future comparison of results with Salway's work, this set of emotion classes provided distinct and easily understandable differentiations between tags. As previously stated, this tag set comprises 21 emotion classes or types which we used as our annotation labels.

Associated with each of these types is a number of descriptive keywords which numbered 627 in total. For the purposes of creating a guide for our annotation process, this list was condensed by removing grammatical variants and semantically very similar words. The intent was to create a succinct and concise list, which would be a practical reference tool for use in a manual annotation task. The finalised tag-set and a list of descriptive terms for each tag are shown in Table 5.1(a).

Emotion Tag	List of Descriptive Terms
Love	Affection, attraction, care for, desire, enchantment, fascination, fondness, friendliness, likeable, preference
Joy	Cheerful, delight, elation, enraptured, exhilaration, festive, glad, happy, mirthful, pleasure
Pride	Arrogance, boastful, conceit, dignity, egotism, exultant, gallant, glorious, haughty, majestic, stuck up, triumphant, vanity
Admiration	Approval, awe, respect
Hope	Anticipation, encouragement, excitement, expectant, optimistic, promise, wishful
Happy For	Pleased for
Relief	Reprieve
Satisfaction	Appreciated, cheering, comforting, complacency, fulfilment
Gratitude	Appreciative, thankful
Gratification	Indulgence, pleased with themselves, smug
Gloat	Revel
Pity	Commiseration, compassion, mercy, misfortune, pathetic, poor, sympathy
Remorse	Guilt, regret, repentance, sorrow
Shame	Disgrace, embarrassment, scandalous
Reproach	Appalled, blame, contempt, disapproval, reprimand
Resentment	Envy, hostility, jealousy, sulkiness
Disappointment	Defeat, failure, heartbroken, hopeless

Distress	Bad, depressed, grief, hurt, pain, pressure, sadness, shock, stressed, suffering, tearful, torture
Anger	Annoyance, exasperation, huffiness, madness, offended
Hate	Disgust, dislike, estrangement, unfriendliness
Fear	Concern, cowardice, cringing, desperate, dismay, dread, horror, hysteria, nervous, panic, reverence, tension, thrill, timid, uneasy, worry

Table 5.1(a): Emotion Tag Set Used for Annotation

5.2 THE ANNOTATION PROCESS

The manual annotation process detailed in this section was used to annotate the selection of feature films described in the next section. For the purposes of this work, portions of each film containing high musical content were of most interest. The means by which these music-rich segments were identified is discussed in the next section. Following this is a step-by-step description of the actual process employed.

5.2.1 Music-Rich Movie Segments

One of the aims of this work is in examining the emotive influence of film music and its enhancing effect on the emotional impact of filmic events. As such, in each film under scrutiny, the sections that were of high interest are those containing high musical content on the movie soundtrack. To that end, each feature film chosen to undergo manual annotation of its emotional content was analysed using the speech-music discrimination process [Lehane, 2006] described in the previous chapter. This process consisted of a movie soundtrack undergoing low-level audio feature analysis and automatic classification at a per second granularity. This resulted in each second of a feature film being classified into five categories, namely *Silence*, *Speech* (dialogue), *Music*, *Quiet Music* and the catch-all, *Other*. The data was then aggregated to a per minute granularity and minutes containing a majority of *Music* and *Quiet Music* were flagged as music-rich sections.

5.2.2 The Manual Annotation Process

All manual annotation was carried out by a single person, namely the author, in order to lend a degree of consistency across the annotation results. It was conducted using a PC equipped with video-playback software catering for the Microsoft AVI format and an earphone headset. The BS.Player^{TM} multimedia player [Webteh Ltd, 2006] was chosen for its keyboard shortcut functionality and the provision of extensive skip-forward and -backward capabilities. The set of 21 emotion tags and their associated descriptions (see Table 5.1(a)) were printed out on paper and mounted on the monitor, to allow for quick referral during the movie.

Each film was annotated on a per-minute granularity and the aim was to allow the annotator to apply one or more emotion labels to each minute of the film in an efficient manner. To that end a spreadsheet application was used to design an interactive, colour-formatted annotation matrix. As can be seen in Figure 5.2(a) above, the spreadsheet has a column for each minute in the film, and a row for each of the 21 emotion labels. In each cell a number could be entered, where the digit "1" was the indicator for an emotion present in film scenes contained in that minute. One or more emotions could be entered per minute. On entering the number the cell applied colour-coded conditional formatting, with the colour indicating the associated emotion label.

The Process

1. A blank annotation matrix file was created from the template, and adjusted to the appropriate length for the particular film.
2. The music-rich segments were calculated from the speech-music discrimination results for the particular film, and then entered into the matrix file. These segments were indicated in the matrix by changing the background colour of the cells in the minute-columns of the music-rich segments to green and grey (see Figure 5.2(a) for examples). Also indicated is the "frequency of occurrence" column, which displays a running total of the annotations currently made for each emotion.
3. Once the spreadsheet was prepared, the movie file was opened and the desktop windows were arranged as shown in Figure 5.2(b), with the spreadsheet set to show the first column.

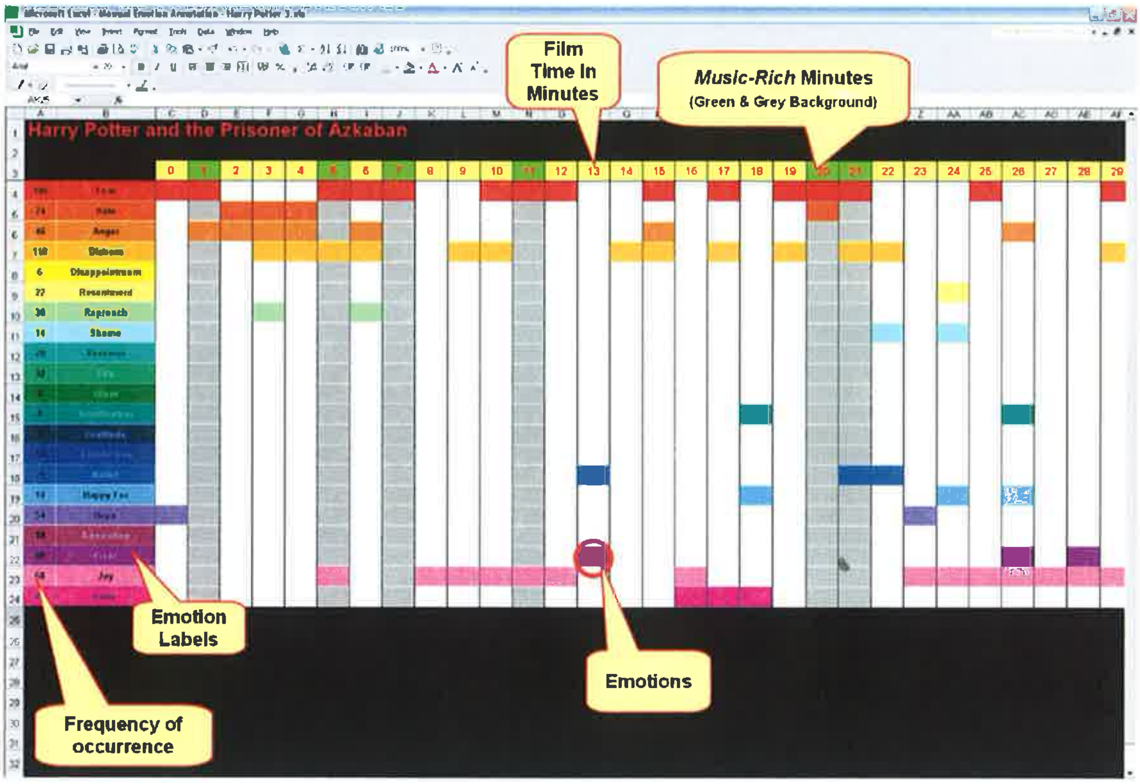


Figure 5.2(a): Labelled Screen Shot of the Movie Annotation Spreadsheet Matrix

4. Once the film started playing, the annotator simply had to observe what was happening in the film and while watching the film in any minute logging the emotion or emotions present in that minute in the appropriate column in the spreadsheet.
5. If required the film could be stopped and a scene replayed if necessary.
6. The annotator was required to take a break from annotation every 20 to 30 minutes, depending on the film action, for at least 5 to 10 minutes. This was to ensure that a high level of concentration could be maintained constantly throughout the annotation process. Particular attention was to be paid to those music-rich minutes.
7. At the end of the film, when the credits started rolling, the annotator stopped annotating and marked the current minute column as containing the start of the end credits.

The results of this annotation process are given in Section 5.4 below.

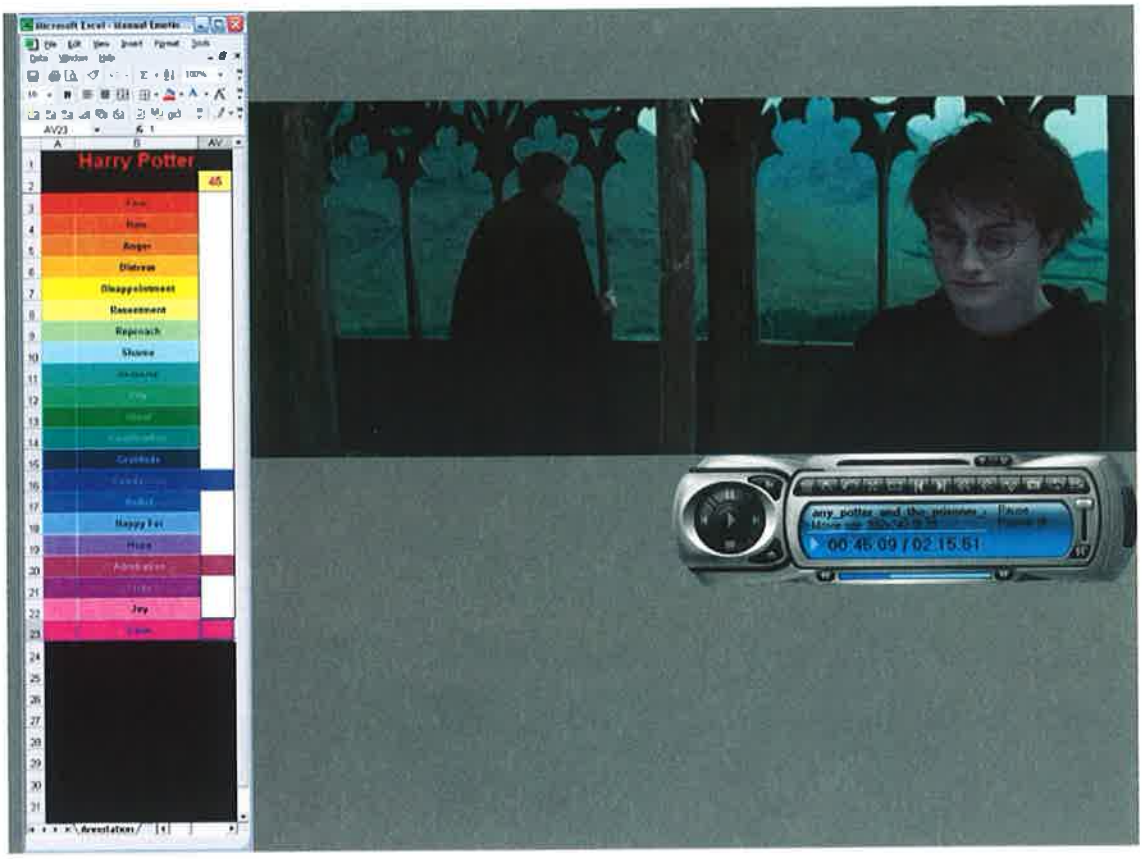


Figure 5.2(b): Screen Shot of a Movie Annotation in Progress

5.3 MANUALLY ANNOTATED FEATURE FILMS

5.3.1 Movie Genres

As was discussed in Chapter 3, a total of fifty-one feature films were shown in the CDV*Plex* experiment, encompassing 10 film genres. In the participant surveys, viewers were asked to rate this list of genres in order of preference with the results shown below in Figure 5.3(a).

As can be seen from this chart, the top five most popular genres overall were comedy, thriller/mystery, action/adventure, romance and science-fiction/fantasy, with animated movies running a close sixth. Comic films were a favourite among both genders while, perhaps somewhat predictably, female viewers showed a definite preference for romantic movies while male viewers showed more of a preference for action/adventure, horror/suspense and science-fiction/fantasy. More unusually there was also a definite male preference for both foreign films and documentaries.

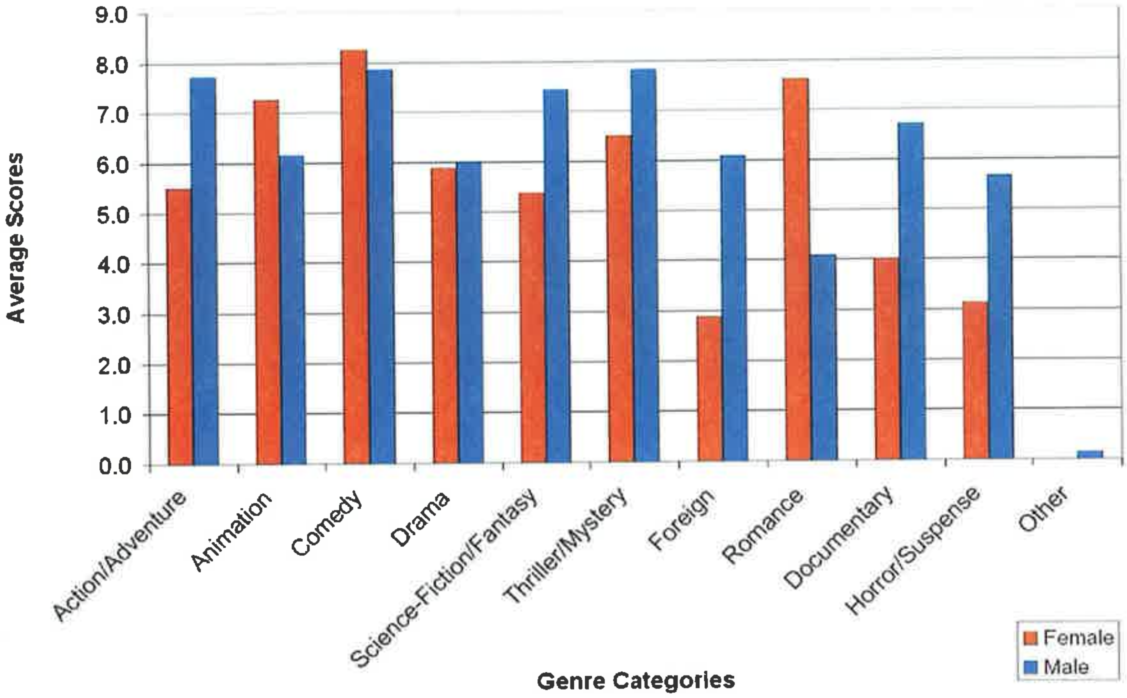


Figure 5.3(a): Most Popular Film Genre Categories as per CDV*Plex* Participants

5.3.2 Feature Film Selection

When choosing films for manual annotation, it was important to select a range of films which reflected the preferences of viewers and therefore films were chosen from the across the top five most popular genres as indicated above. The motivation for this being that if viewers enjoy a certain type of film more, they are more likely to become emotionally involved in the story of that film and hence their reactions should be stronger, which should allow for more successful correlation analysis with the physiological data gathered in the CDV*Plex* experiment, as outlined in Chapter 3.

The selection of films chosen to undergo manual annotation includes films from the 1940s, 1970s, 1980s, as well as more modern choices, giving a good spread across several different film-making eras. This is in addition to film genre considerations previously outlined. The films chosen for manual annotation are listed below in Table

5.3(a). It should be noted that the film chosen from the "comedy" genre is also an animated film, thus overlapping the genre of animation which was sixth in viewers' preference. Further information on each of the six films, including detailed synopses, can be found in Appendix D.

Film Genre	Manually Annotated Film
Comedy/Animation	Finding Nemo
Thriller/Mystery	Mean Streets
Action/Adventure	Pirates of the Caribbean: The Curse of the Black Pearl Indiana Jones and the Last Crusade *(aka. Indiana Jones III)*
Romance	Casablanca
Sci-Fi/Fantasy	Harry Potter and the Prisoner of Azkaban *(aka. Harry Potter III)*

Table 5.3(a): Films chosen for manual annotation by genre

5.4 ANNOTATION RESULTS

This section includes a brief summary of the annotation results, before detailing the results of each annotation on a per film basis.

5.4.1 Music Detection

As discussed in Section 5.2.1 each annotated film also underwent low-level audio feature analysis and automatic classification, as described in Chapter 4.3.1, in order to detect the presence of speech and music on the movie soundtrack. This ultimately resulted in each minute of a movie being classified into five categories, namely *Silence, Speech, Music, Quiet Music* and *Other*. A breakdown of the results of this speech-music discrimination, with respect to occurrences of music, is given below in Table 5.4(a) and illustrated in Figure 5.4(a).

This summary allows us to see the actual number of minutes within each movie in which music was detected. For example, we can observe that *Indiana Jones III* has the highest proportion of music content at 77 minutes, which is to say 64% of the movie has a detectable musical presence.

Film	Length (hh:mm)	Instances of Speech-Music Discrimination (mins)			Total Music (mins)	Music Rich Proportion of Film
		Music	Quiet Music	Speech and Music		
Finding Nemo	01:34	9	7	0	16	17%
Mean Streets	01:45	24	3	0	27	26%
Indiana Jones and the Last Crusade (III)	02:00	56	21	0	77	64%
Pirates of the Caribbean: The Curse of the Black Pearl	02:15	45	17	0	62	46%
Harry Potter and the Prisoner of Azkaban (III)	02:14	9	34	0	43	32%
Casablanca	01:37	15	7	0	22	23%

Table 5.4(a): Music-Rich minutes in Manually Annotated Films

5.4.2 Emotion Annotation Results

In this section we give an overview of the results of the emotion annotation, and the results are then further discussed on a per film basis below. A legend is included in Figure 5.4(a) showing what colours indicate each emotion type in some of the illustrative charts shown below.

Figure 5.4(a): Legend for Emotion Types as depicted in Annotation Results

Overview

The diagram shown below in Figure 5.4(b) gives a high level view of the average number of occurrences of each type of emotion which were annotated over all six feature films under analysis. From this diagram we can clearly see that *FEAR* is by far the most prevalent emotion type, with *DISTRESS*, *JOY*, *ANGER*, *HOPE* and *LOVE* following in a relatively close cluster.

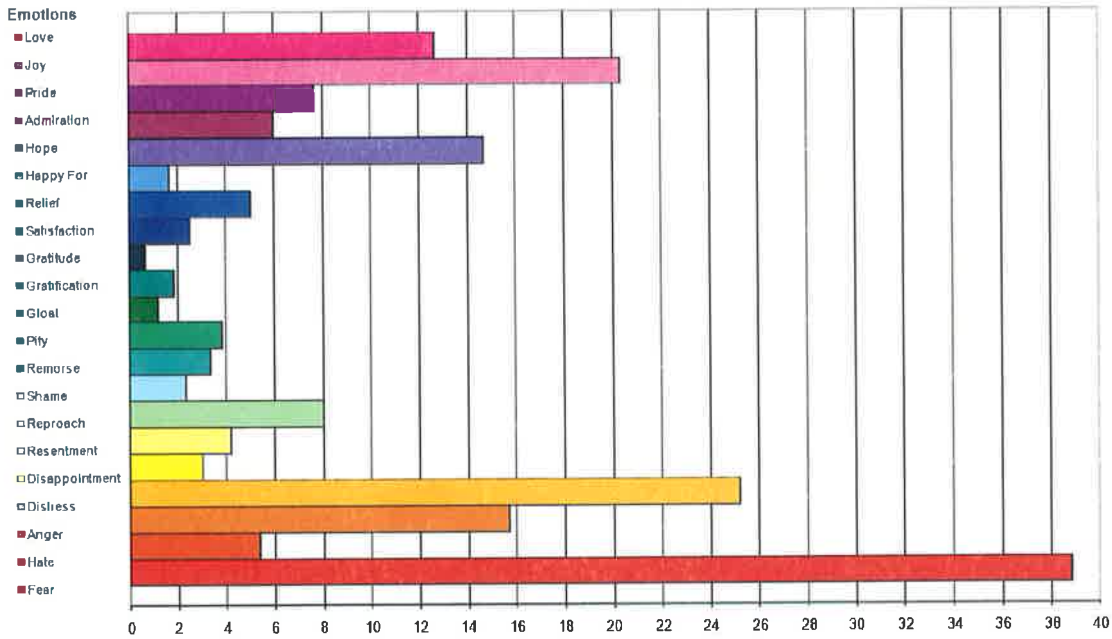


Figure 5.4(b): Average Occurrence of Emotions over all Manually Annotated Films

The following six diagrams are the actual annotation results, a sequential time series representing the length of a film colour coded by emotion type occurrence. As we discussed in Section 4.2.2, each film was annotated for emotion on a per-minute basis, therefore in the diagrams below the numbers ranging across the top of the diagram represent each minute of the movie. When read vertically, each column indicates the combination of emotions which were evident in that minute, based on the colour legend illustrated in Figure 5.4(a). This allows us to visually distinguish patterns and blocks of emotion within the movie. The subsequent six diagrams again show the annotation results, but these also indicate the music-rich portions of each movie. As was previously demonstrated in Figure 5.2(a), if a minute of a movie has been deemed to contain music, as per Section 5.4.1 above, then it's background colour is changed to grey.

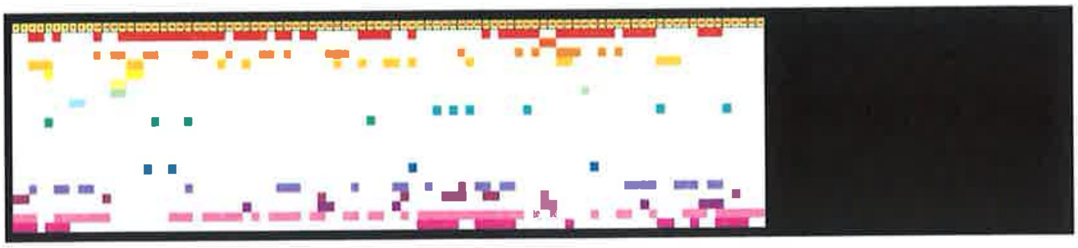


Figure 5.4(c): Distribution of emotions in *Finding Nemo*

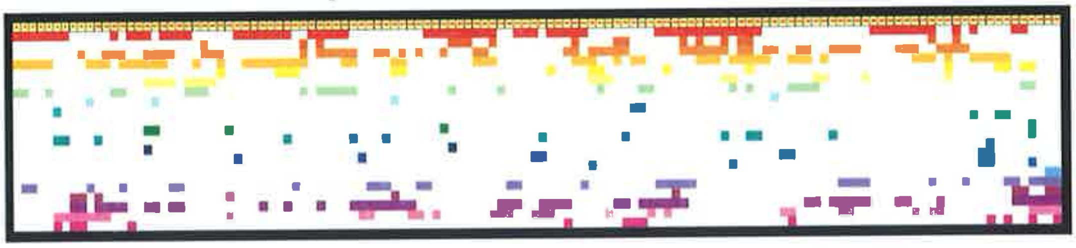


Figure 5.4(d): Distribution of emotions in *Pirates of the Caribbean*

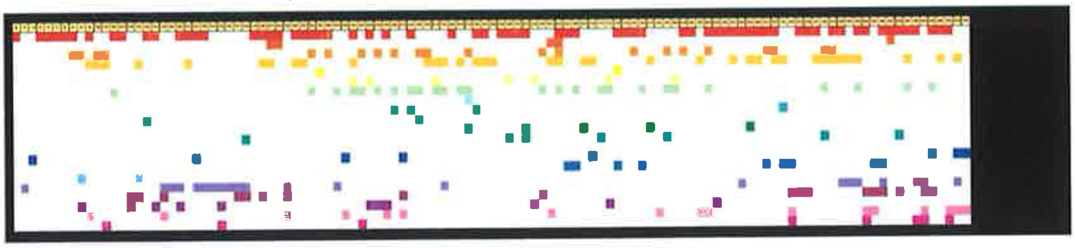


Figure 5.4(e): Distribution of emotions in *Indiana Jones and the Last Crusade*

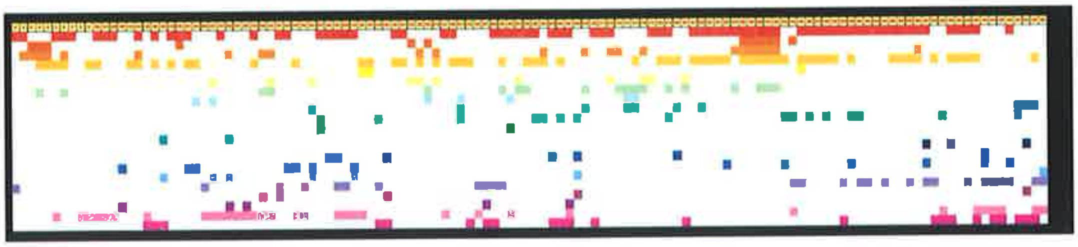


Figure 5.4(f): Distribution of emotions in *Harry Potter and the Prisoner of Azkaban*

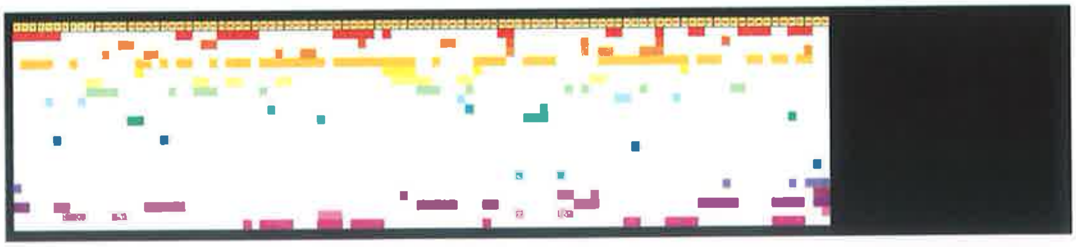


Figure 5.4(g): Distribution of emotions in *Casablanca*

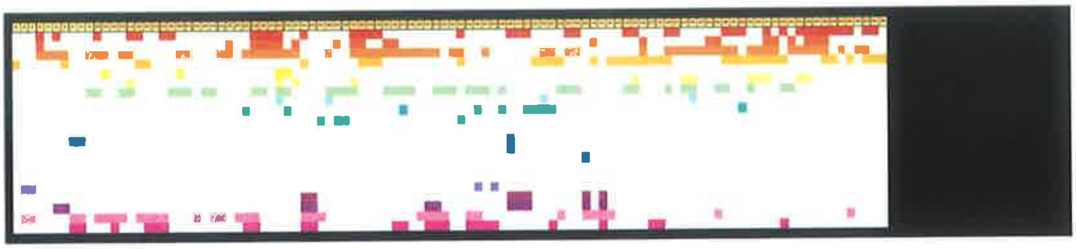


Figure 5.4(h): Distribution of emotions in *Mean Streets*

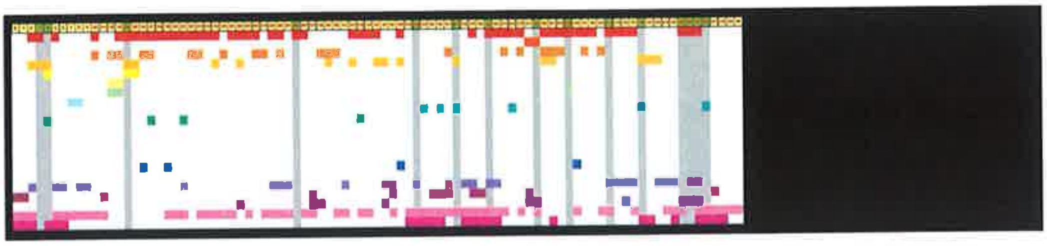


Figure 5.4(i): Distribution of emotions, with Music-Rich Sections indicated, in *Finding Nemo*

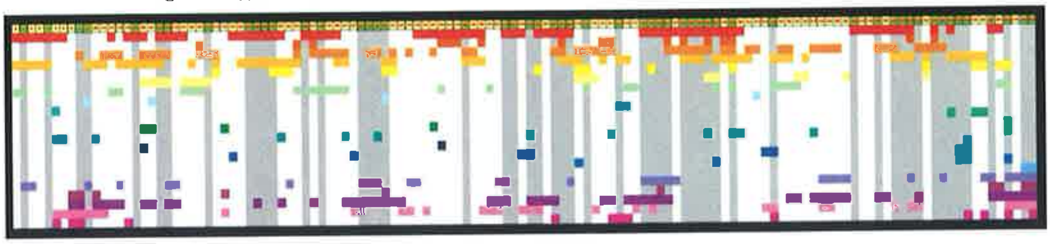


Figure 5.4(j): Distribution of emotions, with Music-Rich Sections indicated, in *Pirates of the Caribbean*

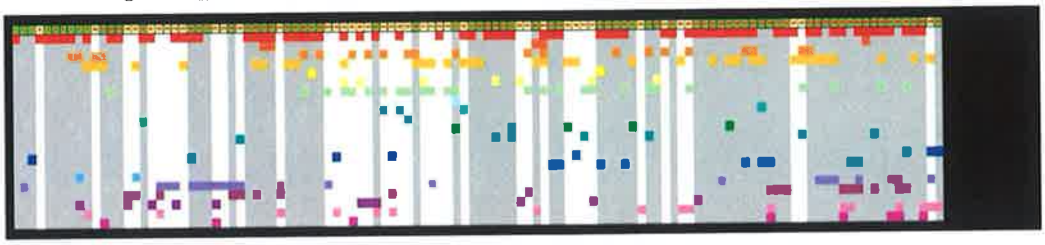


Figure 5.4(k): Distribution of emotions, with Music-Rich Sections indicated, in *Indiana Jones and the Last Crusade*

Figure 5.4(l): Distribution of emotions, with Music-Rich Sections indicated, in *Harry Potter and the Prisoner of Azkaban*

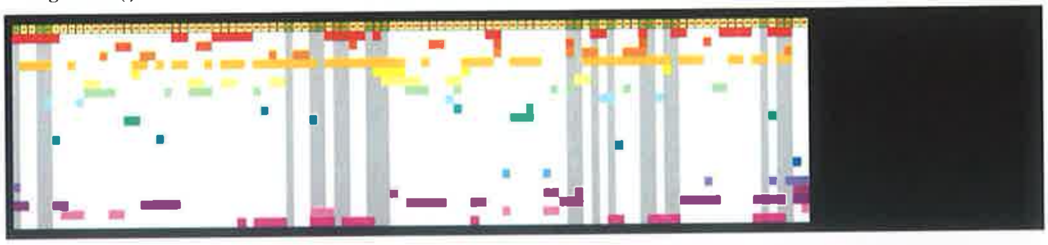


Figure 5.4(m): Distribution of emotions, with Music-Rich Sections indicated, in *Casablanca*

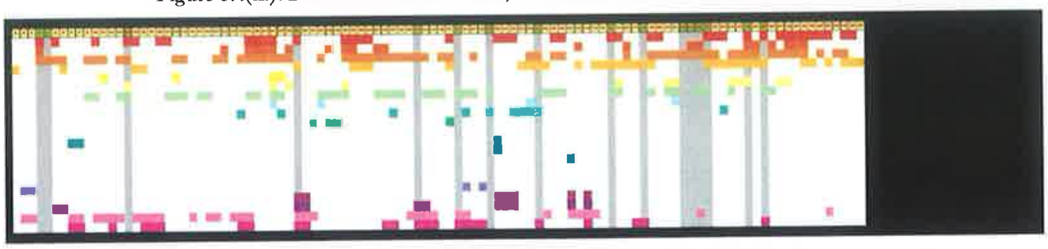


Figure 5.4(n): Distribution of emotions, with Music-Rich Sections indicated, in *Mean Streets*

When examining these annotation results, bear in mind that the red-orange-yellow colours represent the 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 dominancy of these two sets of emotion, which can be regarded as the "opposite extremes" of emotion, is clearly evident in the results shown above, where we consistently see temporally long blocks of these emotion types frequently occurring. Referring to Table 5.1(a) a broader sense of what each emotion represents can be gleaned from its associated description of related terms.

Finding Nemo

The annotation results for *Finding Nemo* are shown in Figure 5.4(c), where we can observe the temporal distribution of emotion type occurrences. *Finding Nemo* is the shortest of all six films chosen for annotation at just 94 minutes. As one would expect of a film in the comedy genre there are significant blocks of *JOY* prevalent throughout the film, and as can be seen below in Figure 5.4(o) it is the dominant emotion. Occurrence of *FEAR* is also high, which when considered in terms of some of its descriptive words, "concern", "worry", "panic" and "nervous", is not surprising with respect to the film's plot, which focuses on a father's search for his missing son, animated characters of talking tropical fish notwithstanding. *HOPE, LOVE, ANGER* and *DISTRESS* also have similar and relatively high frequencies of occurrence.

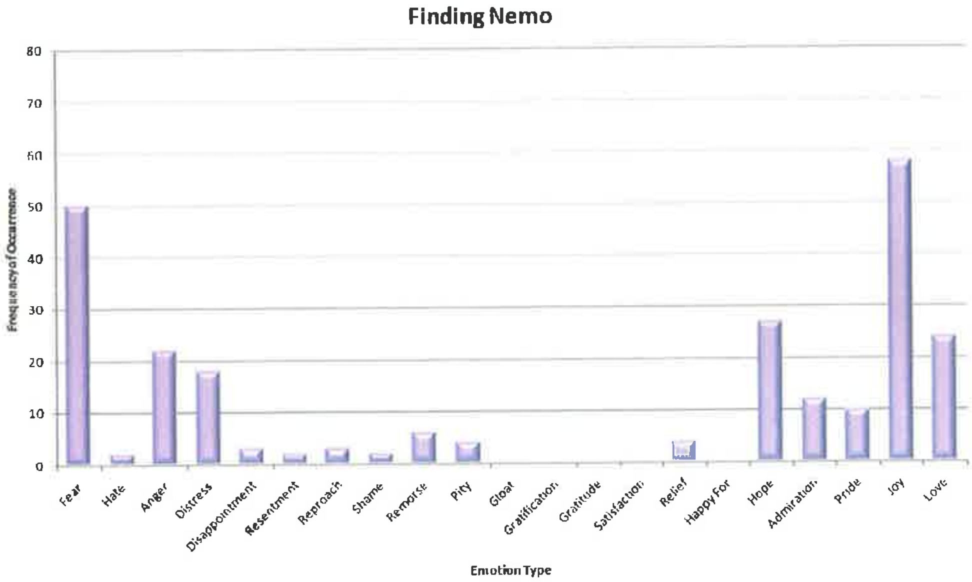


Figure 5.4(o): Frequency of emotions in *Finding Nemo*

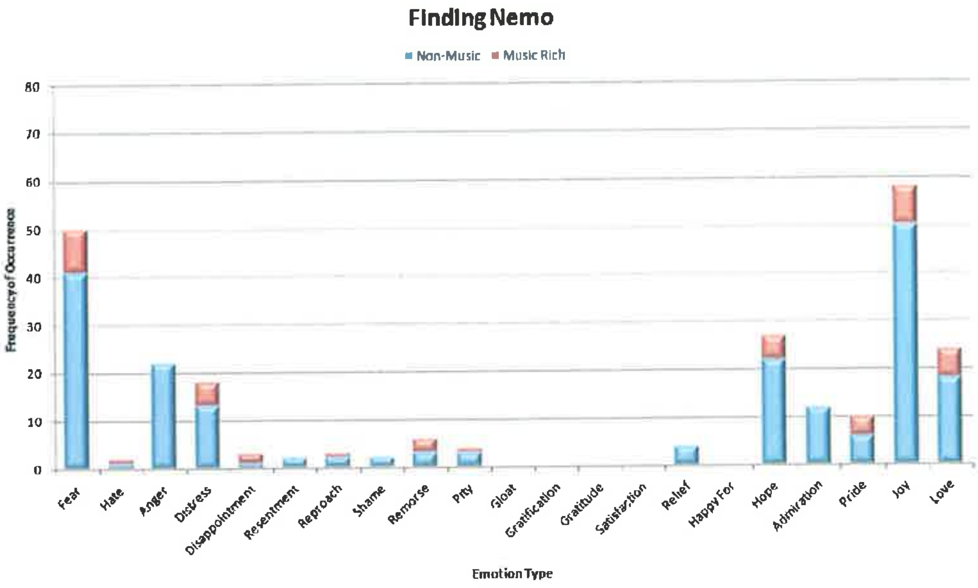


Figure 5.4(p): Frequency of emotions w.r.t. Music-Rich sections in *Finding Nemo*

Finding Nemo has a relatively low rate of music detection, with just 17% of the movie considered to be music-rich, illustrated in Figure 5.4(i). The proportion of the occurrences of each emotion annotated as occurring in the music-rich segments of the film is indicated in Figure 5.4(p) above.

Pirates of the Caribbean: The Curse of the Black Pearl

The annotation results for *Pirates of the Caribbean* are shown in Figure 5.4(d), where we can observe the temporal distribution of emotion type occurrences. *Pirates of the Caribbean* is the longest of all six films chosen for annotation at 135 minutes. Unsurprisingly for a film in the action/adventure genre *FEAR, ANGER* and *DISTRESS* are present in significant blocks throughout the film, and as can be seen below in Figure 5.4(q) these are the dominant emotions. Occurrences of *PRIDE* are also high, which when considered in terms of some of its descriptive words, "arrogance", "exultant", "gallant" and "egotism", is not surprising with respect to the film's plot, which is a period swashbuckling tale of swaggering pirates and romantic heroes. *JOY* and *HOPE* also have similar, relatively high frequencies of occurrence reflecting the comedy and excitement of this old-fashioned pirate tale with a modern twist.

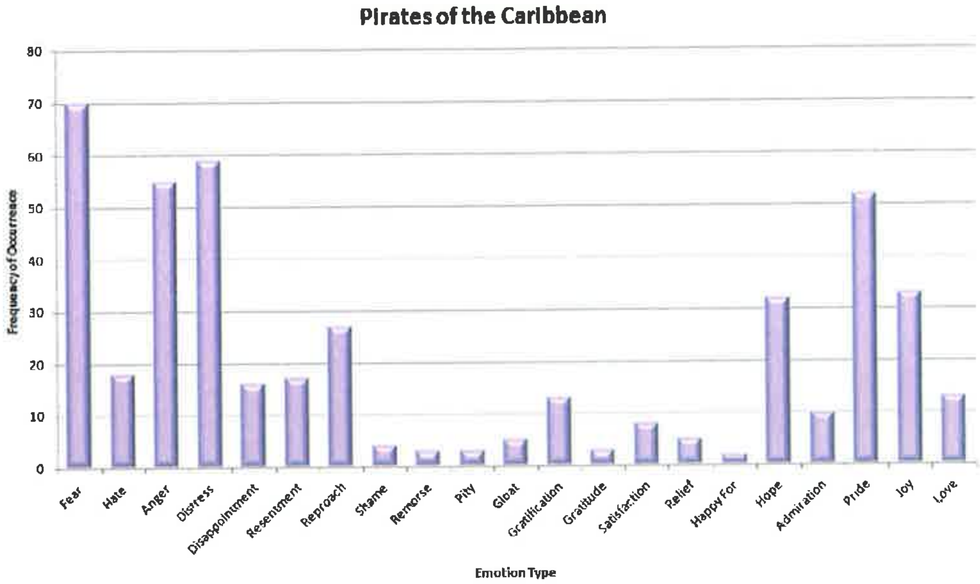


Figure 5.4(q): Frequency of emotions in *Pirates of the Caribbean*

Pirates of the Caribbean has a very high rate of music detection, with 46% of the movie considered to be music-rich, illustrated in Figure 5.4(j). The proportion of the occurrences of each emotion annotated as occurring in the music-rich segments of the film is indicated in Figure 5.4(r), where we observe that a large proportion of the occurrences of each dominant emotion are music-rich. For example, half of the instances of *FEAR* occur with music present.

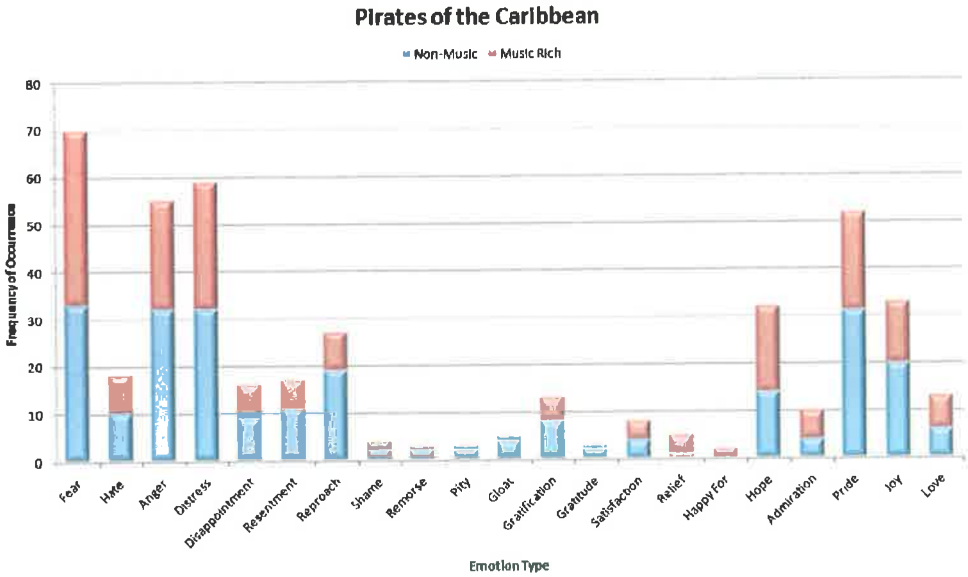


Figure 5.4(r): Frequency of emotions w.r.t. Music-Rich sections in *Pirates of the Caribbean*

Indiana Jones and the Last Crusade

The annotation results for *Indiana Jones and the Last Crusade (aka. Indiana Jones III)* are shown in Figure 5.4(e), where we can observe the temporal distribution of emotion type occurrences. *Indiana Jones III* is quite long at 120 minutes. Once again for a film in the action/adventure genre the emotions of *FEAR, ANGER* and *DISTRESS* are present in significant blocks throughout the film, and as can be seen below in Figure 5.4(s) these are, once again, dominant emotion. Occurrences of *REPROACH* are also high, which when considered in terms of some of its descriptive words, "blame", "disapproval" and "reprimand", is not surprising with respect to the film's plot, in which an archaeologist adventurer rescues his estranged and crotchety father from the Nazis. *HOPE* and *JOY* also have similar, relatively high frequencies of occurrence reflecting the comedy and excitement inherent in this classic adventure story.

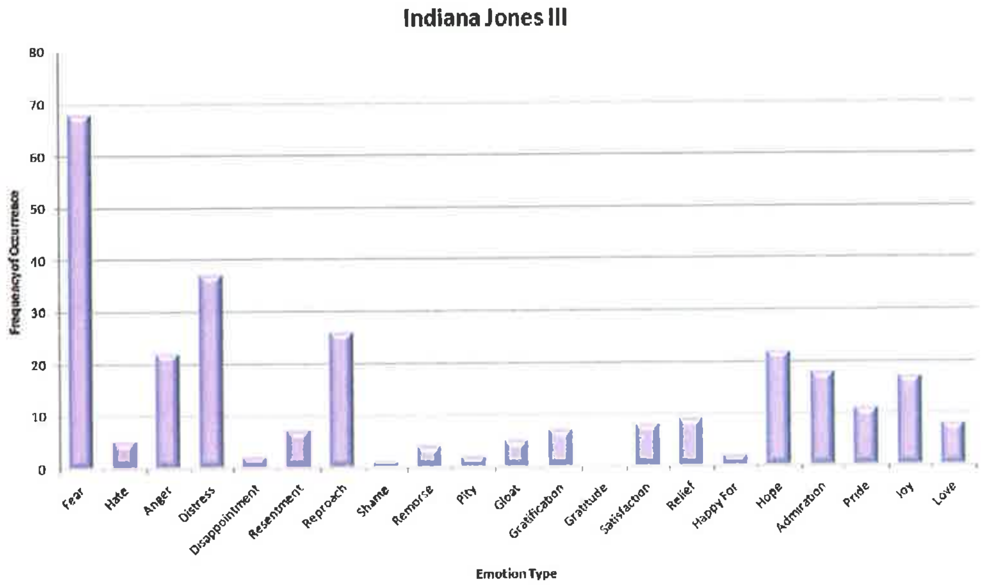


Figure 5.4(s): Frequency of emotions in *Indiana Jones and the Last Crusade*

Indiana Jones III has the highest rate of music detection, with 64% of the movie considered to be music-rich, illustrated in Figure 5.4(k), which is a typical feature of the classic Hollywood adventure. The proportion of the occurrences of each emotion annotated as occurring in the music-rich segments of the film is indicated in Figure 5.4(t), where we observe that, for most emotion types, the majority of instances of emotion occur in music-rich segments. For example, the vast majority of instances of *FEAR* occur with music present.

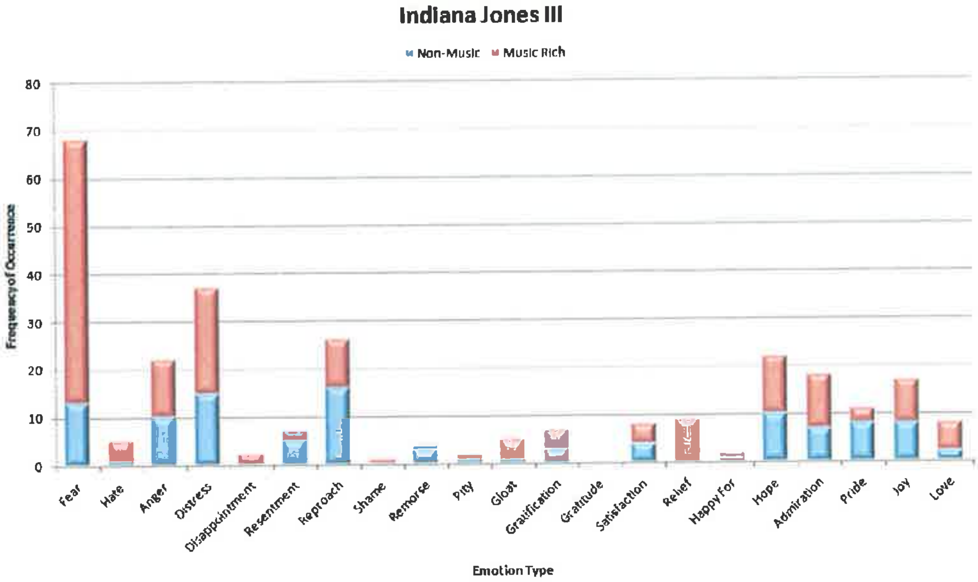


Figure 5.4(t): Frequency of emotions w.r.t. Music-Rich sections in *Indiana Jones and the Last Crusade*

Harry Potter and the Prisoner of Azkaban

The annotation results for *Harry Potter and the Prisoner of Azkaban (aka. Harry Potter III)* are shown in Figure 5.4(f), where we can observe the temporal distribution of emotion type occurrences. *Harry Potter III* is the second longest of all six films chosen for annotation at 134 minutes – just 1 minute shorter than *Pirates of the Caribbean*. Though primarily a children's fantasy tale of magic and wizardry, this is also quite dark tale of evil wizards, horrific and deadly creatures and a deranged murderer on the run. Given the plot, it is unsurprising that emotions of *FEAR* and *DISTRESS* are present in significant blocks throughout the film, and as can be seen below in Figure 5.4(u) these are the dominant emotions by far. However there are considerable elements of comedy and affection as the plot centres around a group of kids in magic school, reflected in the high proportions of *HOPE, JOY* and *LOVE*. Occurrences of *ANGER* and *HATE* are also high, which is not surprising when the main character's battles with the creatures responsible for his parents' murderers and also with horrendous relatives and teachers.

Harry Potter III has a high rate of music detection, with 32% of the movie considered to be music-rich, illustrated in Figure 5.4(l). The proportion of the occurrences of each emotion annotated as occurring in the music-rich segments of the film is indicated in Figure 5.4(v), where we observe that a significant proportion of the occurrences of the dominant emotion are music-rich. For example, over a third of the instances of *FEAR* occur with music present.

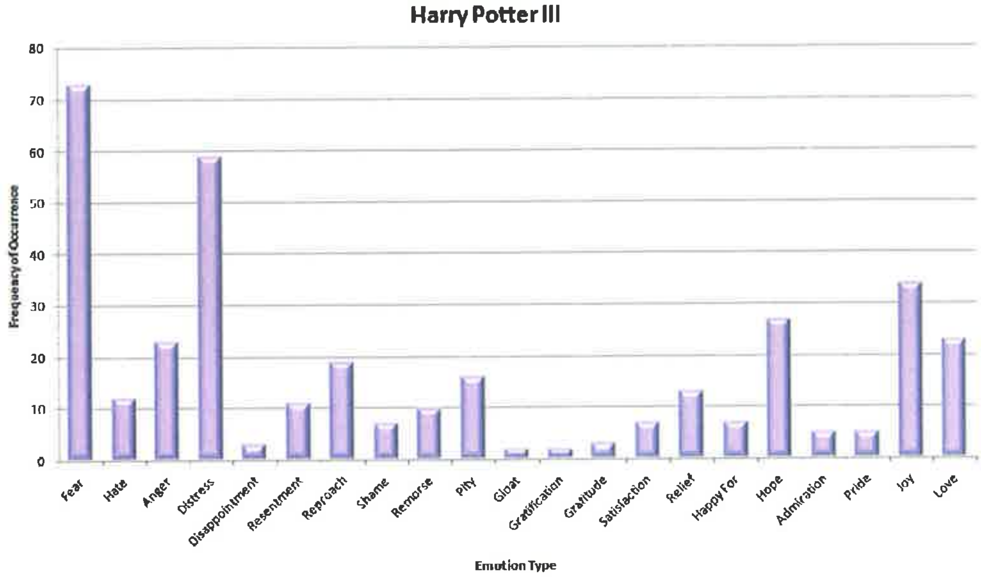


Figure 5.4(u): Frequency of emotions in *Harry Potter III*

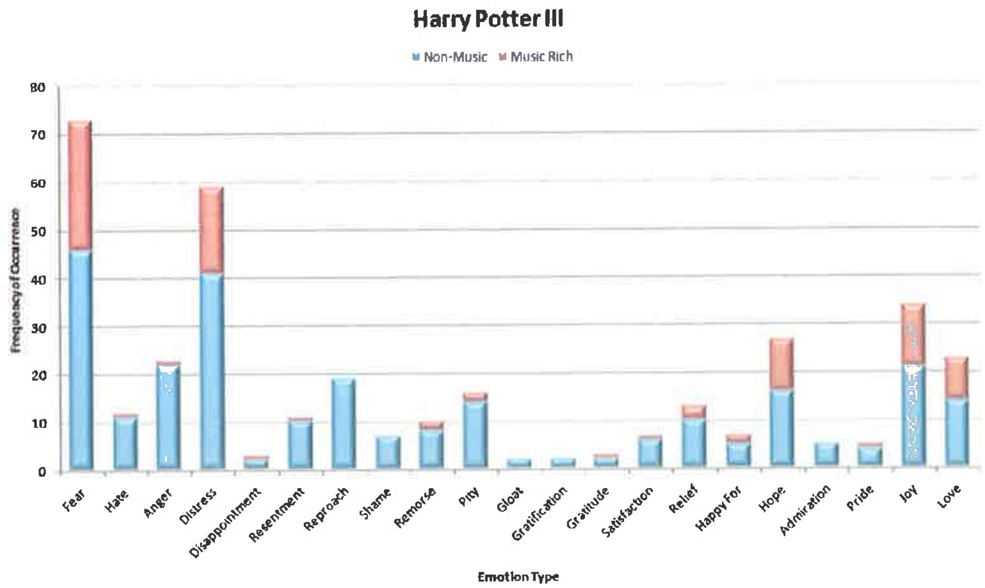


Figure 5.4(v): Frequency of emotions w.r.t. Music-Rich sections in *Harry Potter III*

Casablanca

The annotation results for *Casablanca* are shown in Figure 5.4(g), where we can observe the temporal distribution of emotion type occurrences. *Casablanca* is the second shortest of all six films chosen for annotation at 97 minutes – just 3 minute longer than *Finding Nemo*. It is a classic tale of lost love and emotional turmoil in World War II. Given the plot, it is unsurprising that emotion *DISTRESS* is present in the most significant blocks throughout the film, and as can be seen below in Figure 5.4(w) is the dominant emotion by far. However there are considerable elements of fear and pride as the plot is set in the surroundings of Vichy controlled Casablanca with its strong Nazi presence and involves people desperately trying to escape to America, reflected in the high proportions of *FEAR* and *PRIDE*. Occurrences of *LOVE* are also high, which is not surprising as the main characters tragic story of love found and sacrificed forms the backbone of the plot.

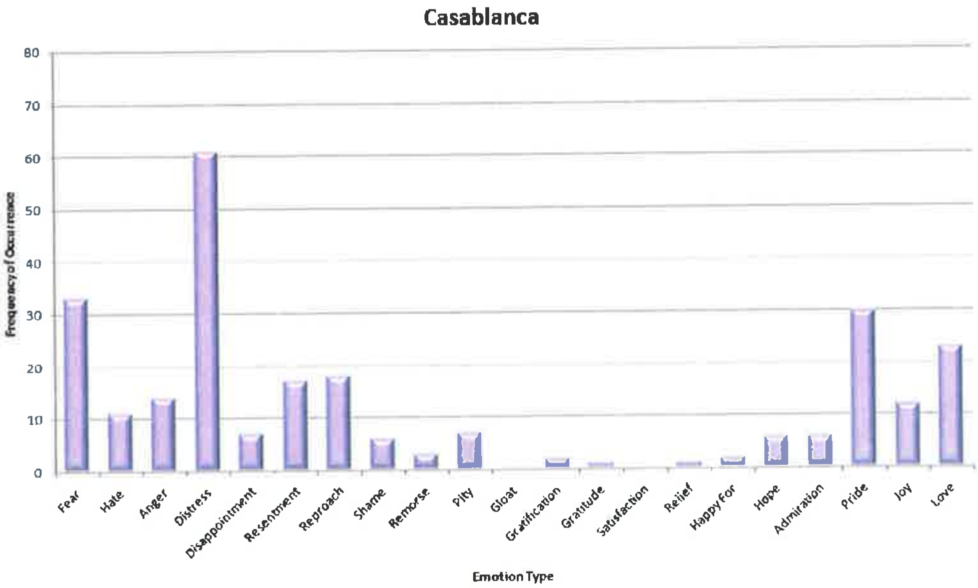


Figure 5.4(w): Frequency of emotions in *Casablanca*

Casablanca has a relatively high rate of music detection, with 32% of the movie considered to be music-rich, illustrated in Figure 5.4(m). The proportion of the occurrences of each emotion annotated as occurring in the music-rich segments of the film is indicated in Figure 5.4(x), where we observe that a relatively significant proportion of the occurrences of the dominant emotion are music-rich. For example, nearly a third of the instances of *DISTRESS* occur with music present.

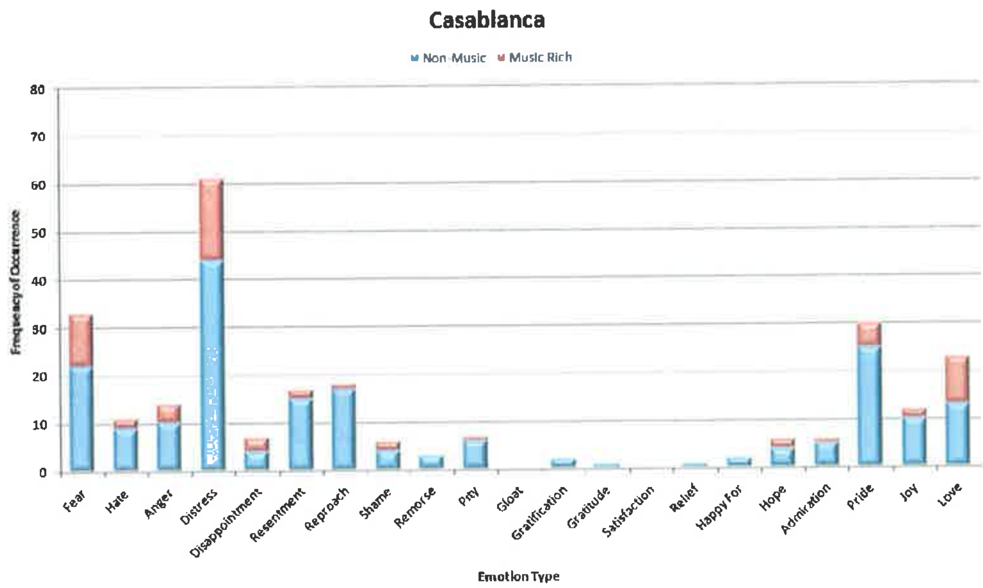


Figure 5.4(x): Frequency of emotions w.r.t. Music-Rich sections in *Casablanca*

Mean Streets

The annotation results for *Mean Streets* are shown in Figure 5.4(h), where we can observe the temporal distribution of emotion type occurrences across its 105 minute length. It is gritty story of crime and violence in 1970s New York, which centres on the conflict experienced by a group of young men as they make their way in life as loan sharks, debt collectors and hooligans. Given the plot, it is unsurprising that emotion *ANGER* is present in the most significant blocks throughout the film, and as can be seen below in Figure 5.4(x) is the dominant emotion by far. The constant conflicts experienced by the main characters are reflected in the high proportions of *REPROACH, FEAR* and *DISTRESS*. Occurrences of *JOY* and *LOVE* are also quite high, up to the last third of the film, where it can be observed to trail off, which is not surprising as the characters relationships and affection for one another all begin to break down and finally implode .

Mean Streets has a relatively low rate of music detection, with 26% of the movie considered to be music-rich, illustrated in Figure 5.4(n). The proportion of the occurrences of each emotion annotated as occurring in the music-rich segments of the film is indicated in Figure 5.4(z).

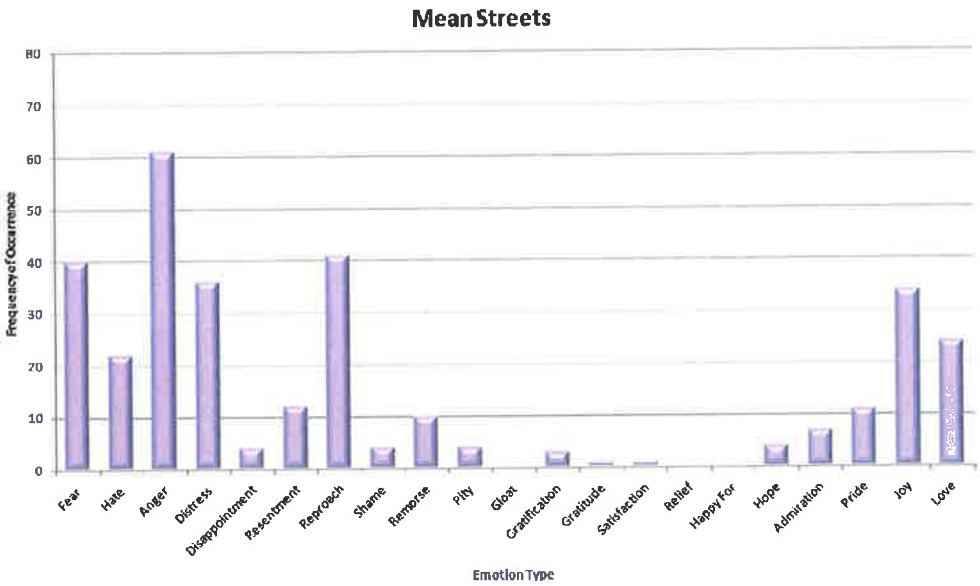


Figure 5.4(y): Frequency of emotions in *Mean Streets*

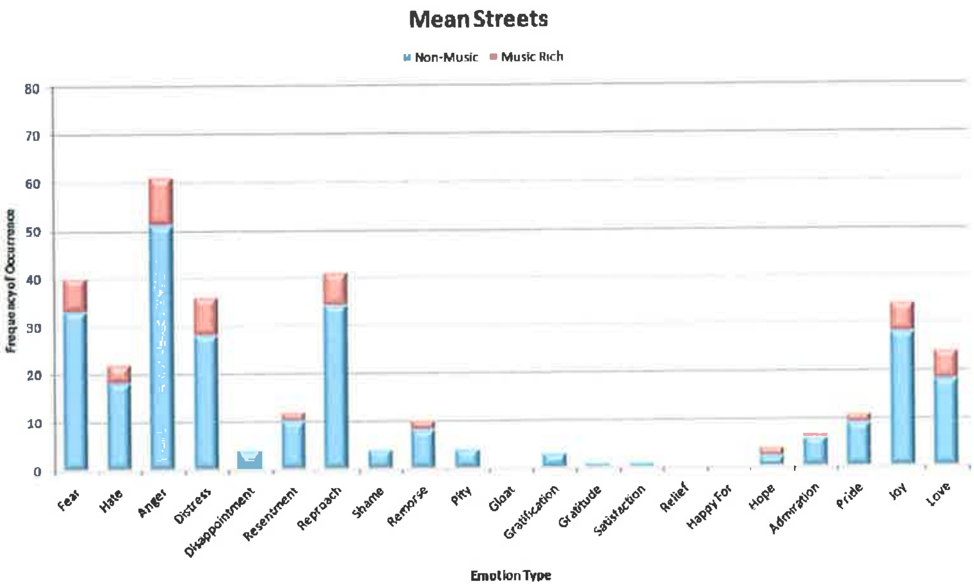


Figure 5.4(z): Frequency of emotions w.r.t. Music-Rich sections in *Mean Streets*

In this chapter we discussed the process used to manually annotate movies for the presence of different types of emotion and we introduced the set of emotion tags used to describe these emotion types. We presented and discussed the annotation results for each of the six films chosen for further analysis in this thesis, including a breakdown of the frequencies and trends of emotion occurrence .

CHAPTER 6

RESULTS AND ANALYSIS

This chapter presents and examines the results of the experiments carried out and discussed in this work and describes our findings. We evaluate the data extracted from the biometric measurements recorded in the CDVPlex experiment and investigate the findings with respect to the three fundamental hypotheses put forward in this work. We analyse the results in terms of the sixteen individual viewings from which they were drawn and the twenty-one emotion types we seek to detect. We investigate the correlations between what was observed in the annotation process and what can be detected from analysis of the biometric data. We then consider the effect of music and events on physiological responses to emotions.

In previous chapters we expounded the three major hypotheses underpinning this work; firstly, that emotional responses experienced by people watching movies can be detected using biometric measurements; secondly, that individual viewers experience similar reactions to emotional stimuli; and finally, that movie music and filmic events in a movie influence viewers' emotional responses.

In order to address these hypotheses we introduced the concept of using feature films in order to investigate how people react to emotional stimuli. We discussed the psychological basis behind this premise, with particular focus on the emotional impact a film's musical score can have on the viewers' perception of emotion in that film. We also discussed how different filmic elements such as the aforementioned film music, as well as action, dialogue and montage events can be detected automatically using audio-visual analysis. We conducted an experiment, known as the CDV*Plex*, with the aim of providing a ground-truth of film viewers' physiological responses for a set of movies. This data was then processed in order to identify "peaks", or areas of interest, within the biometric measurements where the subjects experienced a significant physiological reaction. Once we had pinpointed these biometric "highlights", we then had to consider the type of emotion to which they corresponded. In order to further analyse this data, a defined set of emotion categories was selected, and each film used in our experiment was manually annotated for the presence of these emotion types.

This chapter then, is the culmination of the analysis of all these streams of data – the series of significant physiological responses detected for every viewing of each film; the set of observed emotion categories for each film; the set of automatically detected occurrences of music, and action, dialogue and montage events. We will address the three hypotheses, and discover whether these suppositions are supported by the data.

In the next section, we will consider the data relating to each individual film viewing and analyse it in term of each of the six films viewed and the twenty-one emotion categories used in annotation. We will compare the data obtained for each viewing of a film, and then compare the data aggregated for each film, in order to investigate the second hypothesis which states that individual viewers experience similar reactions to emotional stimuli. The subsequent section then concerns the observed emotional occurrences, as per the annotation process, and the detected biometric peaks, in order to investigate the first hypothesis which states that emotional responses can be detected using biometric measurements. We will examine the emotion category detection rate and compare what was observed in annotation with what was actually biometrically detected. Finally, we will examine the extent to which film music and events affect the emotion detection rate in order to investigate the third hypothesis which states that movie events and movie music influence the emotional response of viewers.

6.1 ANALYSIS OF FILM VIEWINGS

As described in Chapter 3, a "peak" is an event in the biometric readings recorded for a film-viewer which exceeded a certain threshold of significance. From the results of the movie annotation described in Chapter 5, we know that certain emotions occur during a film, mapped to the film's timeline, so that at any point in the film we know which of the twenty-one emotion classes used in this experiment are evident. The point in the timeline at which the peak was detected determines its classification, allowing us to describe a biometric peak in terms of its corresponding emotion. Each occurrence of an emotion in the movie has the potential to elicit an emotionally stimulated physiological reaction in each viewer. Therefore each emotion annotation is regarded as a potential peak. This allows us to calculate what proportion of possible peaks were actually recorded and detected for each of the sixteen individual viewings of all six films by fifteen people, and allows us to describe detected peaks in terms of their temporally equivalent emotion.

6.1.1 Peak Detection and Similarity per Viewing

This section describes the peaks or "highlights" detected in the biometric data for each viewing of each of the six films under analysis, where a "viewing" of a film is the set of biometric readings obtained from a single participant in the CDVP*lex* experiment during the screening of that particular film. As described in Chapter 3, for each viewing the biometric data was collated, synchronised and then underwent the peak detection process. The proportion of potential emotionally stimulated peaks actually biometrically detected in each viewing is illustrated below. The degree of similarity between these levels of biometric reaction for each subject is also described, where similarity is 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, as with *Finding Nemo*, *Pirates of the Caribbean* and *Mean Streets*, for example, the similarity measure was calculated for each combination of viewings and the average similarity is plotted.

Casablanca

Casablanca was viewed by two subjects P14 and P21, both male, and was 98 minutes in length. Figure 6.1(a) describes the proportion of potential emotionally stimulated peaks that were detected in each viewing using its biometric data, and the average peak detection rate for *Casablanca*. The overall average rate of emotion detection was 11%. From a visual study of this diagram it is evident that the two subjects experienced similar levels and patterns of emotional stimulation. Figure 6.1(g) then illustrates the high degree of similarity between the emotional peak detection for both viewings of the film, with a mean similarity measure of 94%.

Harry Potter and the Prisoner of Azkaban

Harry Potter III was viewed by two subjects P17 and P30, both male, and was 124 minutes in length. Figure 6.1(b) describes the proportion of potential emotionally stimulated peaks that were detected in each viewing using its biometric data, and the average peak detection rate for *Harry Potter III*. The overall average rate of emotion detection was also 11%. From a visual study of this diagram it is evident that the two subjects experienced very similar levels and patterns of emotional stimulation. Figure 6.1(h) then illustrates the high degree of similarity between the emotional peak detection for both viewings of the film, with a mean similarity measure of 96%.

Indiana Jones and the Last Crusade

Indiana Jones III was viewed by two subjects P09 and P37, both male, and was 115 minutes in length. Figure 6.1(c) describes the proportion of potential emotionally stimulated peaks that were detected in each viewing using its biometric data, and the average peak detection rate for *Indiana Jones III*. The overall average rate of emotion detection increased slightly to 12%. From a visual study of this diagram it is evident that the two subjects experienced quite similar levels and patterns of emotional stimulation. Figure 6.1(i) then illustrates the high degree of similarity between the emotional peak detection for both viewings of the film, with a mean similarity measure of 91%.

Finding Nemo

Finding Nemo was viewed by four subjects P10-f, P23-f, P28-f and P35-f, all female, and was 90 minutes in length. Figure 6.1(d) describes the proportion of potential emotionally stimulated peaks that were detected in each viewing using its biometric data, and the average peak detection rate for *Finding Nemo*. The overall average rate of emotion detection was once again 11%. From a visual study of this diagram it is evident that the subjects experienced relatively similar levels and patterns of emotional stimulation. As can be seen from Table 6.1(a) below, the subjects' average pair-wise similarity measures are quite close, all within an 8% range, with a mean similarity measure of 91% across all viewings. Subjects P10-f and P35-f are closest, with 95% overall similarity, while P10-f and P28-f are the least similar. Figure 6.1(j) illustrates the degree of similarity between the emotional peak detection for all viewings of the film.

Viewing Comparison Combinations		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%
Average Similarity Measure		**91%**

Table 6.1(a): Similarity values for Viewings of *Finding Nemo*

Pirates of the Caribbean: The Curse of the Black Pearl

Pirates of the Caribbean was viewed by three subjects, P04 and P11, who are both male, and P10-f, who is female, and was 127 minutes in length. Figure 6.1(e) describes the proportion of potential emotionally stimulated peaks that were detected in each viewing using its biometric data, and the average peak detection rate for *Pirates of the Caribbean*. The overall average rate of emotion detection was another slight increase to 13%. From a visual study of this diagram it is evident that the subjects experienced quite similar levels and patterns of emotional stimulation. As can be seen from Table 6.1(b) below, the subjects' average pair-wise similarity measures are close, all within a 5% range, with a mean similarity measure of 93% across all viewings. Figure 6.1(k) illustrates the degree of similarity between the emotional peak detection for all viewings of the film.

Viewing Comparison Combinations		Similarity Measure
P04	P10-f	96%
P10-f	P11	93%
P04	P11	91%
Average Similarity Measure		**93%**

Table 6.1(b): Similarity values for Viewings of *Pirates of the Caribbean*

Mean Streets

Mean Streets was viewed by three subjects P12, P18 and P36, all male, and was 105 minutes in length. Figure 6.1(f) describes the proportion of potential emotionally stimulated peaks that were detected in each viewing using its biometric data, and the average peak detection rate for *Mean Streets*. The overall average rate of emotion detection was another increase at 16%. From a visual study of this diagram it is evident that the subjects, for a large proportion of emotion categories, experienced very similar levels and patterns of emotional stimulation, though for the emotions *PITY*, *GRATIFICATION*, *GRATITUDE* and *SATISFACTION* reaction levels vary significantly. As can be seen from Table 6.1(c) below, however, the subjects' average pair-wise similarity measures are still very close, all within a 3% range, with a mean similarity measure of 89% across all viewings. Figure 6.1(l) illustrates the degree of similarity between the emotional peak detection for all viewings of the film.

Viewing Comparison Combinations		Similarity Measure
P12	P36	91%
P12	P18	89%
P18	P36	88%
Average Similarity Measure		**89%**

Table 6.1(c): Similarity values for Viewings of Mean Streets

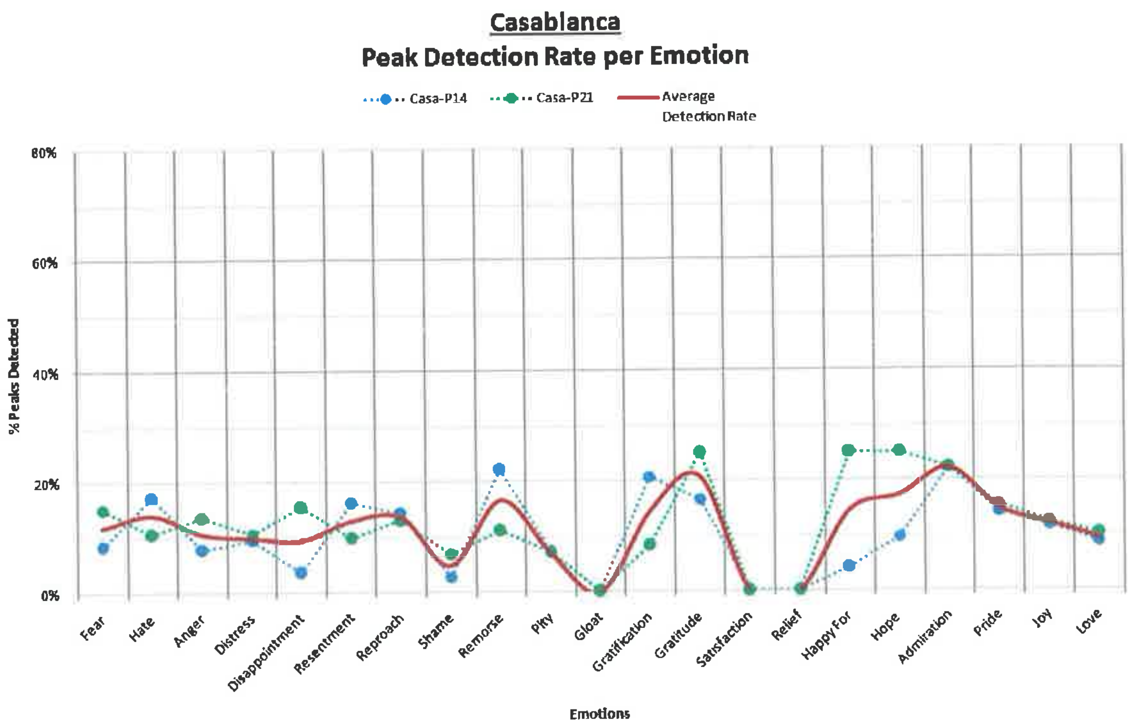


Figure 6.1(a): Emotional Peak Detection per Viewing of *Casablanca*

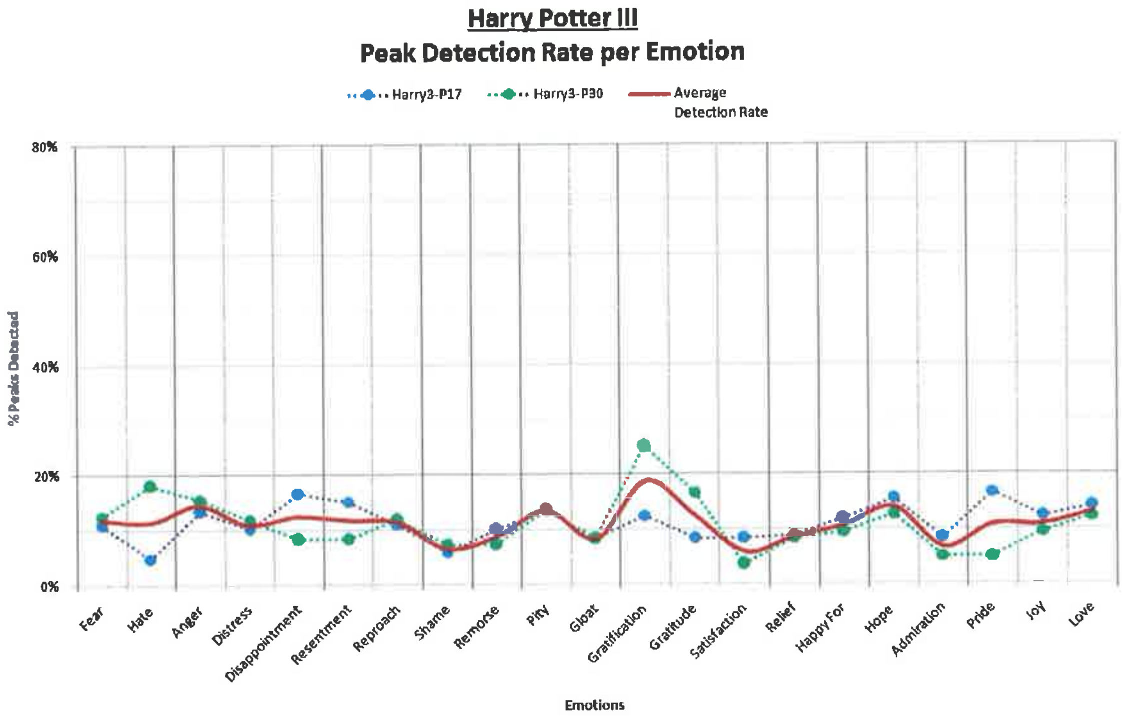


Figure 6.1(b): Emotional Peak Detection per Viewing of *Harry Potter III*

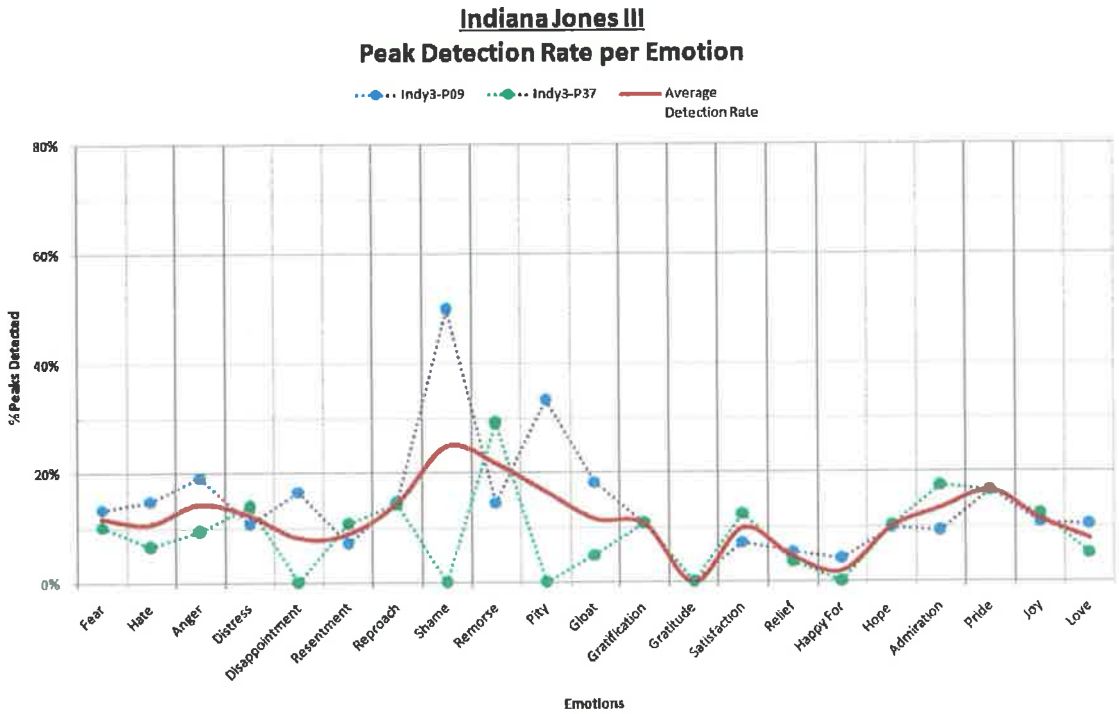


Figure 6.1(c): Emotional Peak Detection per Viewing of *Indiana Jones III*

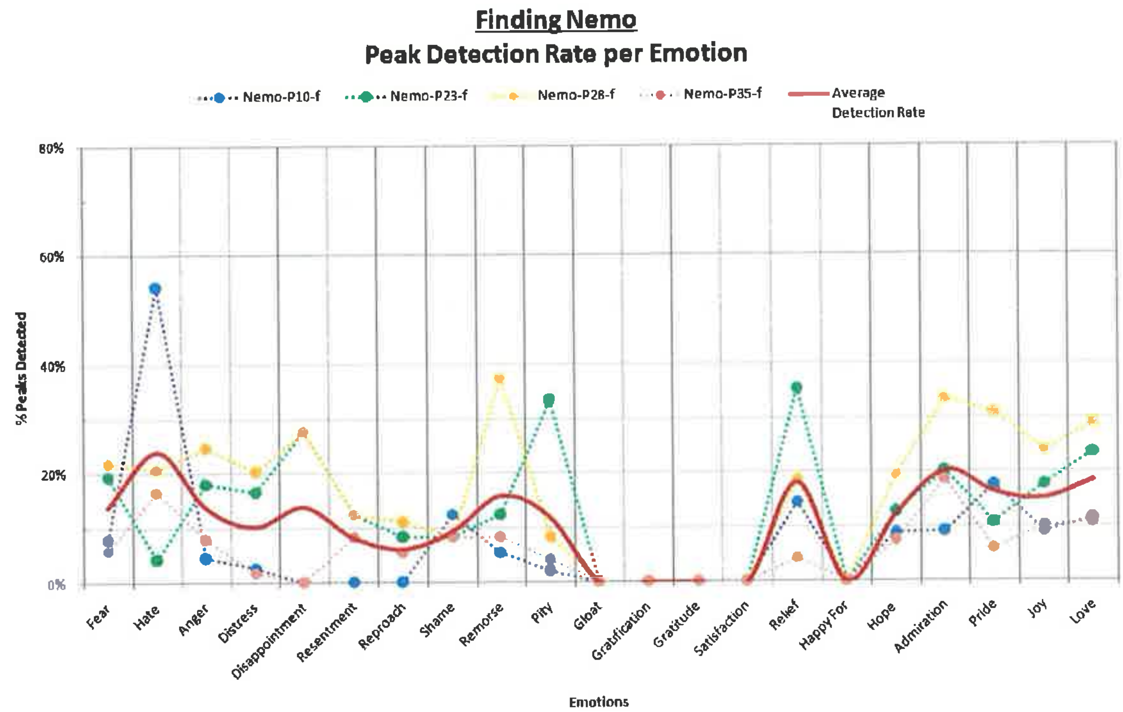


Figure 6.1(d): Emotional Peak Detection per Viewing of *Finding Nemo*

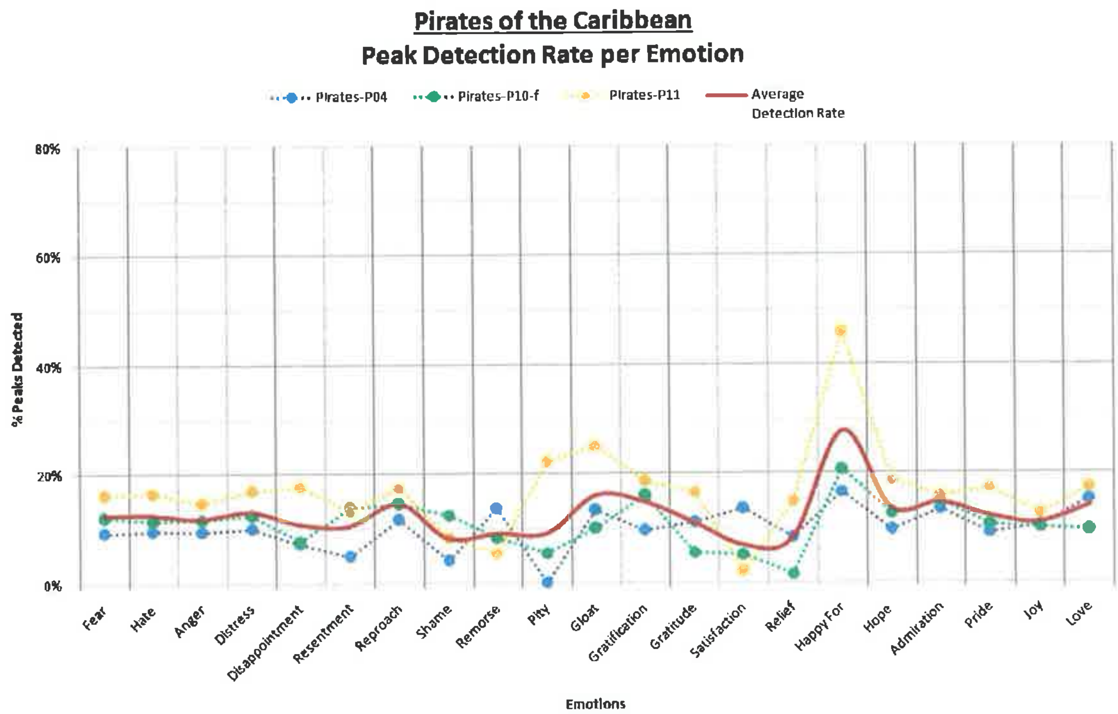


Figure 6.1(e): Emotional Peak Detection per Viewing of *Pirates of the Caribbean*

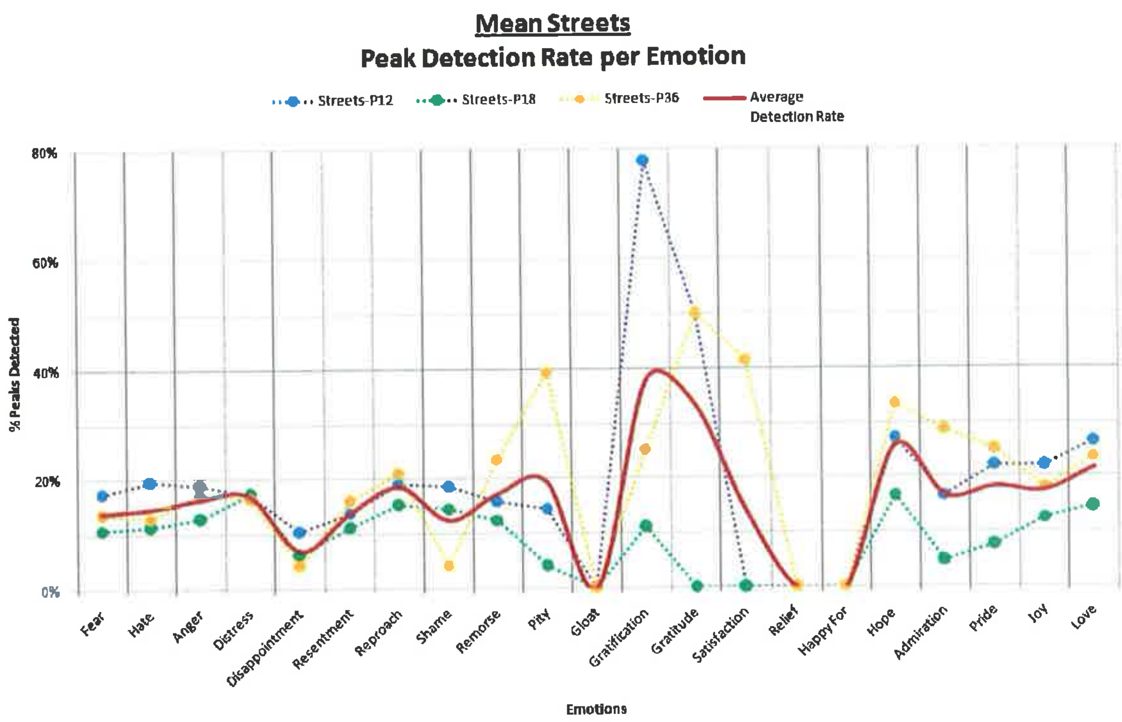


Figure 6.1(f): Emotional Peak Detection per Viewing of *Mean Streets*

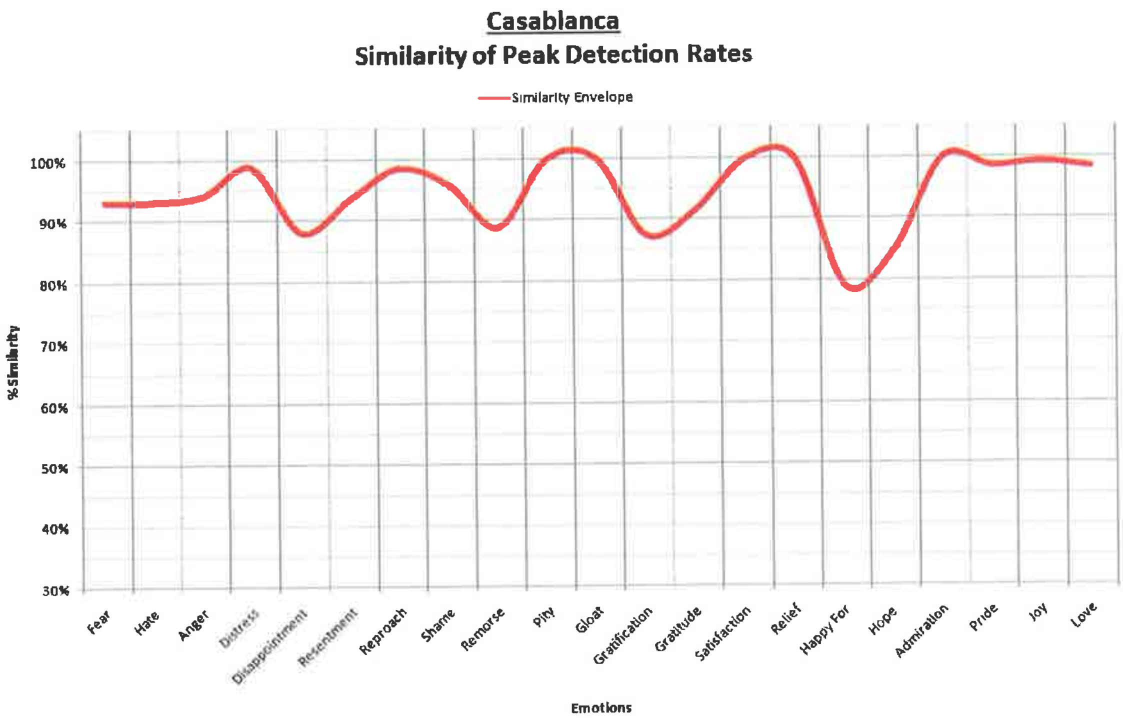


Figure 6.1(g): Similarity of Viewers' Emotional Peak Detection in *Casablanca*

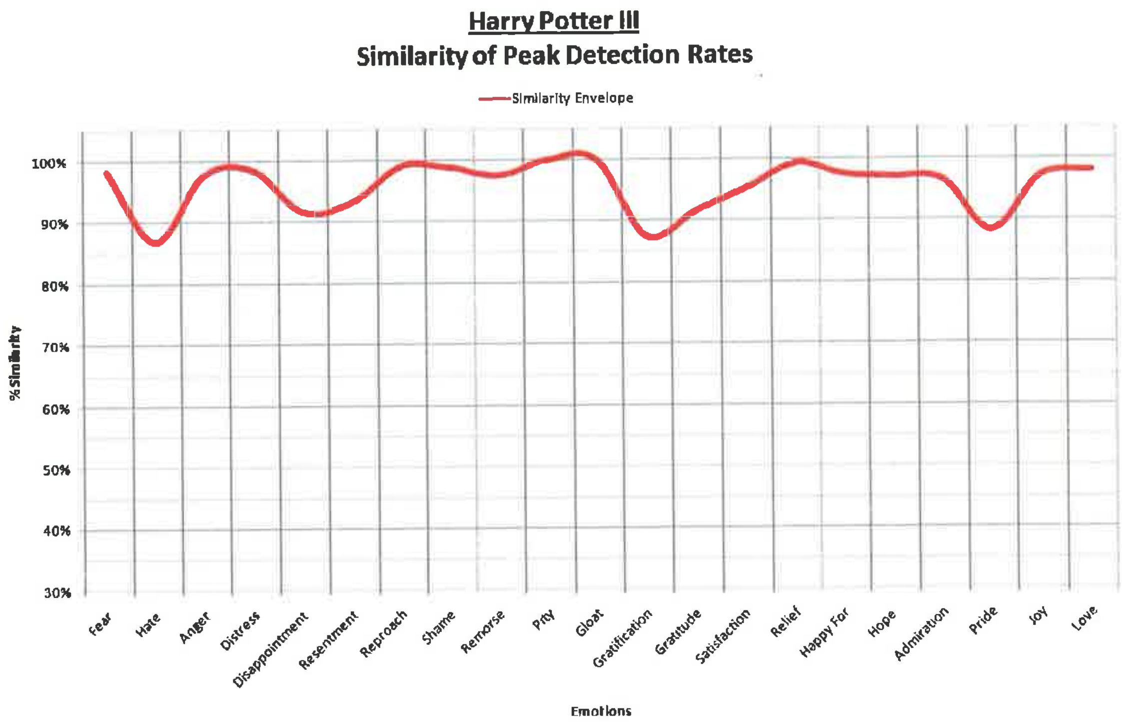


Figure 6.1(h): Similarity of Viewers' Emotional Peak Detection in *Harry Potter III*

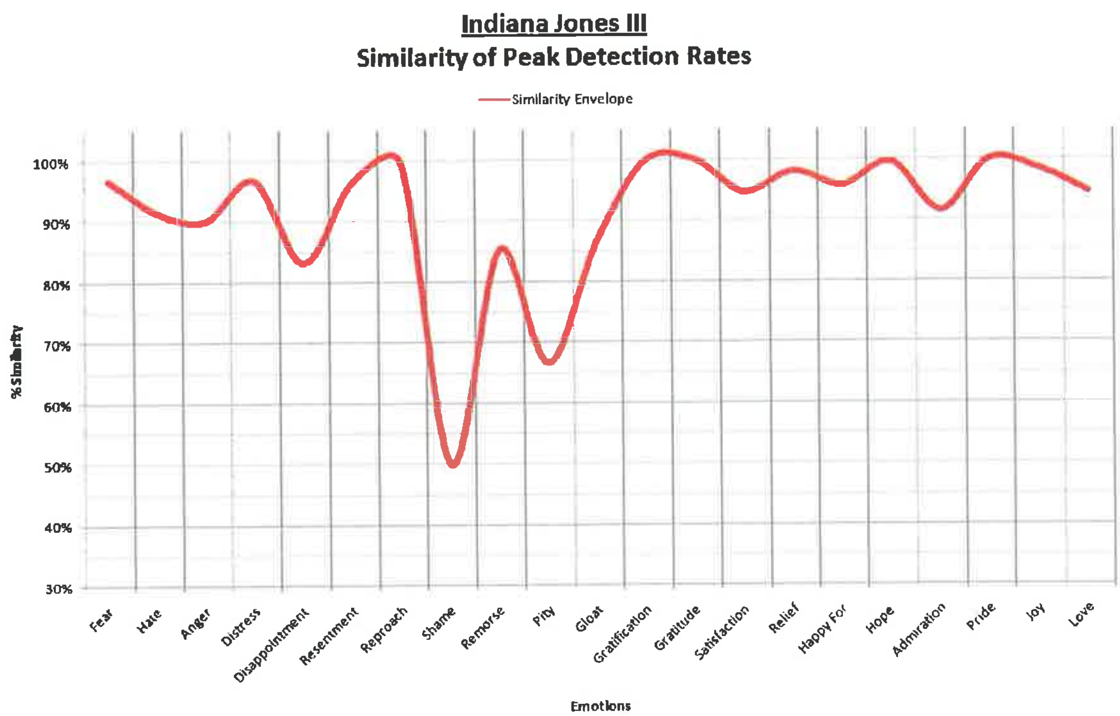


Figure 6.1(i): Similarity of Viewers' Emotional Peak Detection in *Indiana Jones III*

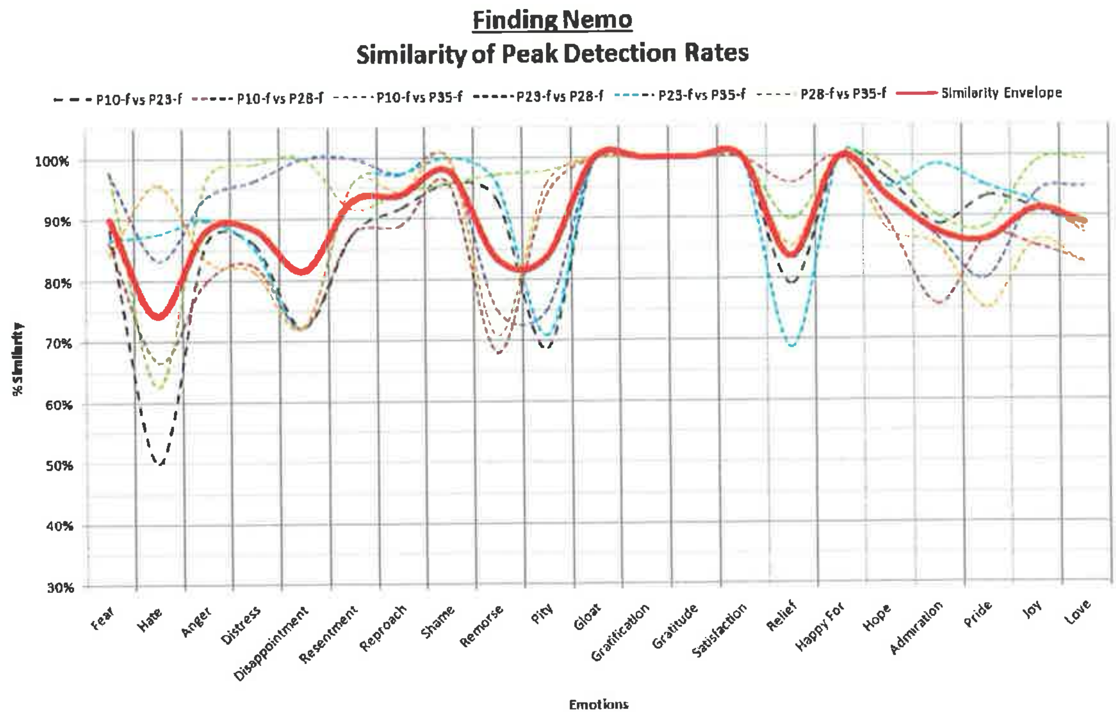


Figure 6.1(j): Similarity of Viewers' Emotional Peak Detection in *Finding Nemo*

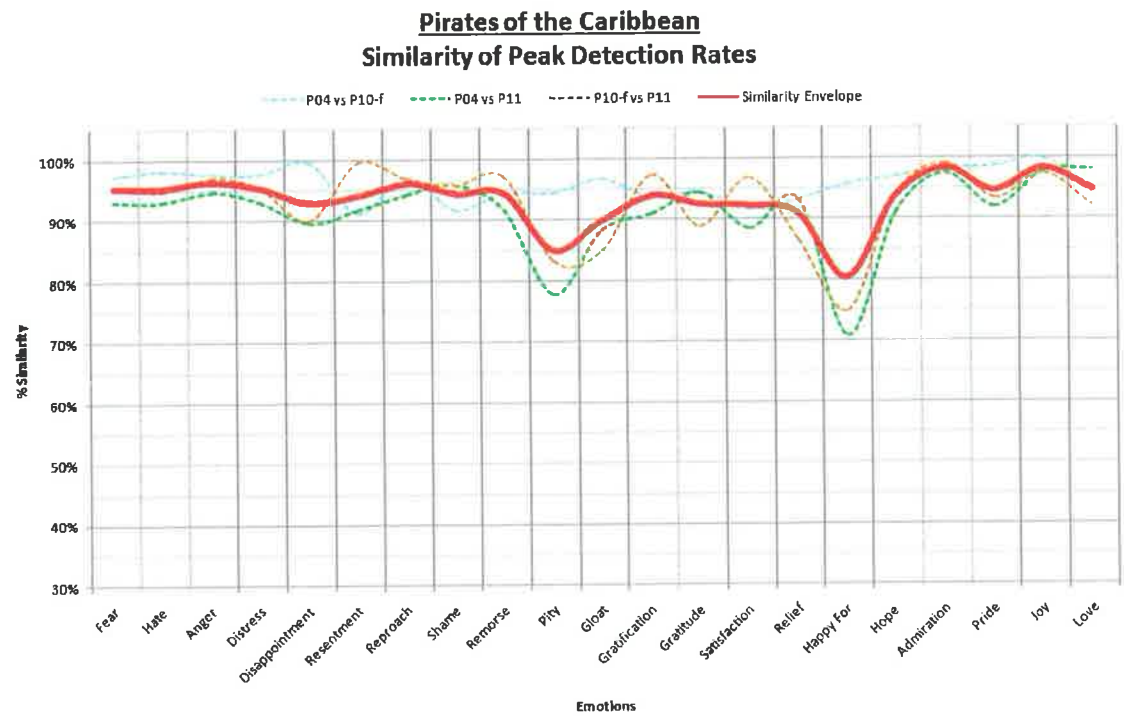


Figure 6.1(k): Similarity of Viewers' Emotional Peak Detection in *Pirates of the Caribbean*

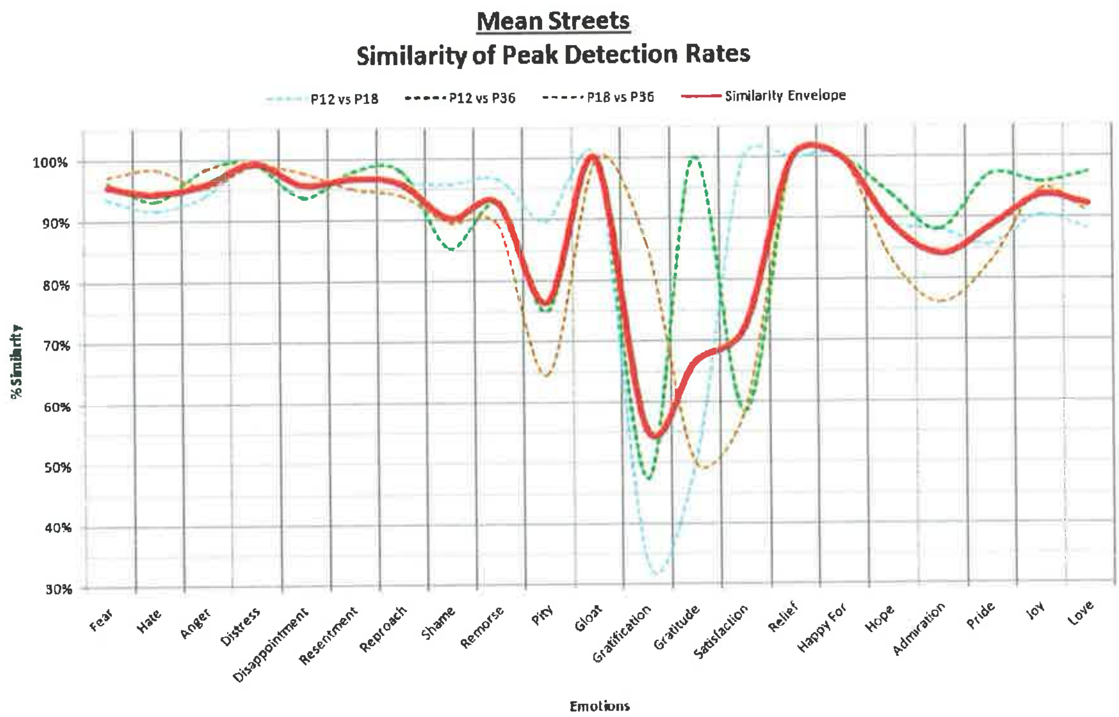


Figure 6.1(l): Similarity of Viewers' Emotional Peak Detection in *Mean Streets*

When taken as a whole, these figures indicate that when people view the same film they respond in a similar manner. The individual rates of peak detection for each film all fall within a 5% range. Also, in all cases but one, the measures of similarity between viewers of the same film were over 90% - in one case as high as 96%. Indeed, the lowest similarity measure that was calculated, for viewings of *Mean Streets,* was still very high at 89%. Table 6.1(d) below gives a brief summary of the results for each film.

Film	No. of Viewers		Mean Peak Detection Rate	Mean Similarity Measure	% Range of Similarity Measures
	Male	Female			
Casablanca	2	0	11%	94%	-
Harry Potter III	2	0	11%	96%	-
Indiana Jones III	2	0	12%	91%	-
Finding Nemo	0	4	11%	91%	8%
Pirates of the Caribbean	2	1	13%	93%	5%
Mean Streets	**3**	**0**	**16%**	**89%**	**3%**

Table 6.1(d): Summary of Peak Detection and Viewer Similarity Results

These results support the premise that individual viewers experience similar reactions to emotional stimuli, where the emotional stimulus is comprised of the same audio-visual material.

6.1.2 Comparison of Viewings across all Movies

While these results, individually, support the second hypothesis put forward in this work, the next step is to investigate the effects of comparing these results across all films. In order to do this, the average peak detection rates and average viewer similarity measures for each film were plotted against each other, and the overall averages across all films were calculated.

Figure 6.1(m), below, illustrates the overall average peak detection rates across all viewings of all films. When we study these results in terms of individual emotions we can see that the peak detection rates are quite close for the most part (for example *ANGER*, *DISAPPOINTMENT* and *RESENTMENT*), with only occasional outliers (for example *GRATIFICATION* in *Mean Streets*, *HATE* in *Finding Nemo*) and some non-response, or zero values (for example *GLOAT* in *Casablanca*, *Finding Nemo* and *Mean Streets*, and *GRATIFICATION*, *GRATITUDE*, *SATISFACTION* and *HAPPY FOR* in *Finding Nemo*).

Figure 6.1(n) illustrates the degree of similarity of emotional peak detection for all viewings of all films. As can be seen from this graph, the overall similarity of responses across all films is high, rising to 95% or above for eight emotion categories and only dipping as far as 85% at its lowest point for the emotion *PITY*.

Overall, we can see that the peak detection rates remain consistent and that similarity measures still remain high, with an overall peak detection rate of 12% and an overall similarity measure of 92% for all viewings of all films. As we have over 95% similarity of responses for nearly 40% of emotions, and an overall average similarity of responses for all emotions of 92%, the results support the hypothesis that viewers experience similar reactions to the same emotions, as in this case the viewers are responding to the emotional stimuli of varied audio-visual material.

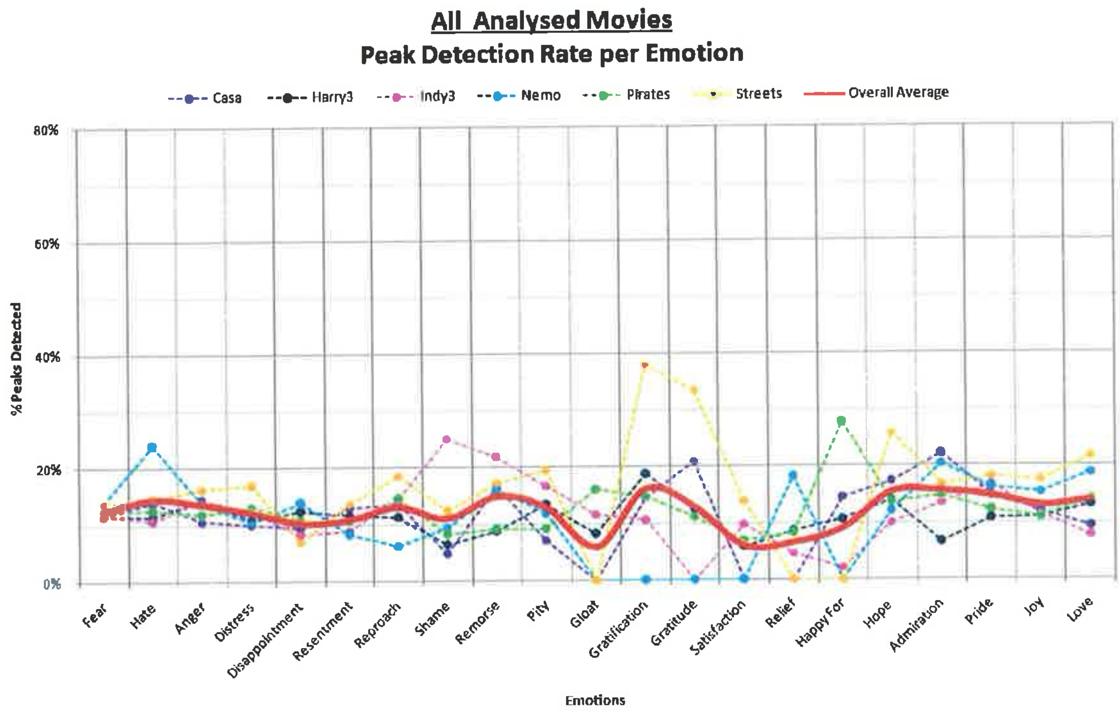


Figure 6.1(m): Overall Emotional Peak Detection across all Movies

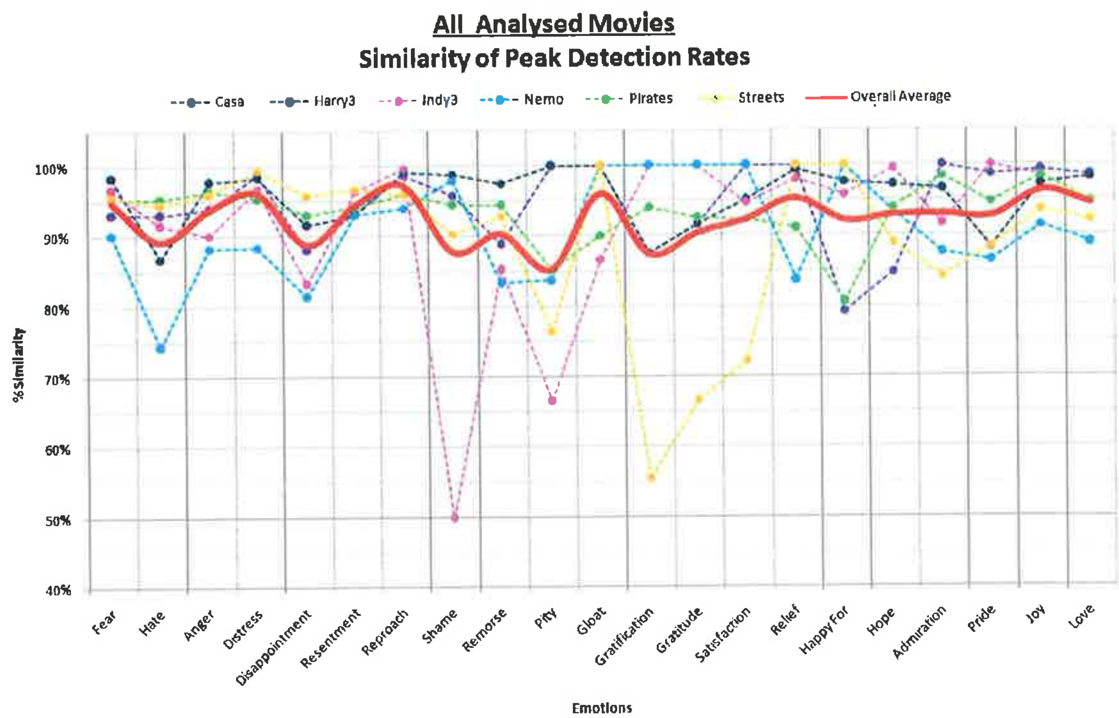


Figure 6.1(n): Overall Similarity of Viewers' Emotional Peak Detection across all Movies

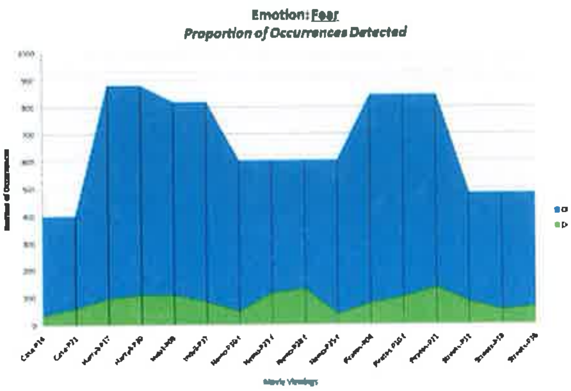
6.2 ANNOTATION VS. DETECTION

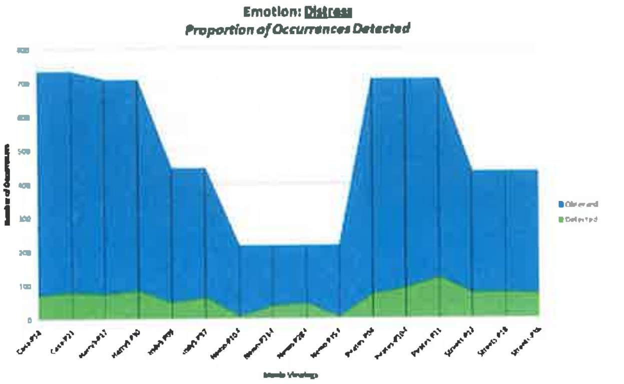
In the previous section, we explored the results of our CDV*Plex* experiment in terms of the movie viewers and the emotion categories. This allowed us to see that the people viewing the films in the biometric cinema all demonstrated a remarkable similarity in how they responded to the emotional stimuli provided by the movies. We also discovered that peak detection rates across these viewings were comparatively low, with an overall average of 12% of peaks detected during the experiment. In this section we will further investigate this phenomenon by expanding our analysis of the emotion annotation, as described in Chapter 5, and the detected peaks in the biometric responses collected in the CDV*Plex* experiment, as described in Chapter 3. We will determine the accuracy of the biometric detection of emotional responses, and therefore address the hypothesis that emotional responses can be detected using biometric measurements.

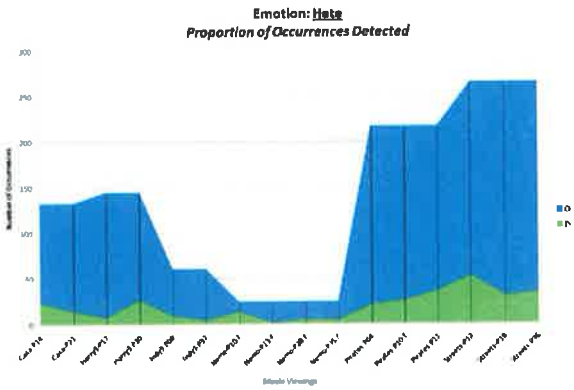
It is worth noting that because of the annotation structure used in this study, whereby the emotion annotations are continuous and therefore no point in an annotated movie will *not* have an emotion associated with it, the focus of this work cannot be to determine whether biometric peaks indicate emotion in the general sense. That is, a peak will *always* have an annotation of an emotion associated with it and therefore it is inaccurate to suggest, under the assumptions made in this work, that the occurrence of a biometric peak can indicate occurrence of emotion, as 100% of peaks detected will correspond with an emotion. The focus of this work, therefore, is to investigate the distribution of the biometric peaks. In this way, we can address the first hypothesis put forward in this work, by examining the detection of emotion in terms of the *type* of emotion. Therefore determining which emotions cause people to react, and to what degree. To that end, in the next section, will revisit the emotion detection rate with respect to each of the twenty-one emotion categories.

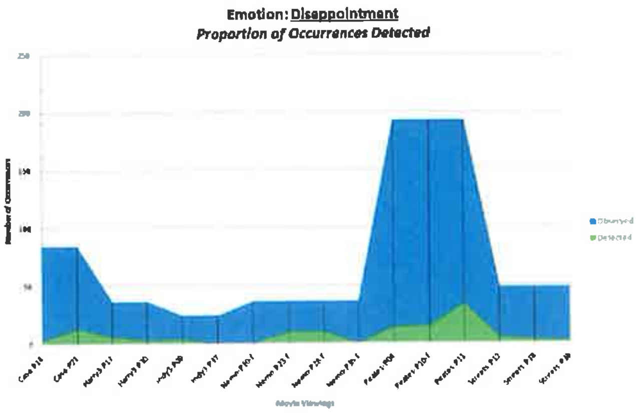
6.2.1 Detection Rate per Emotion

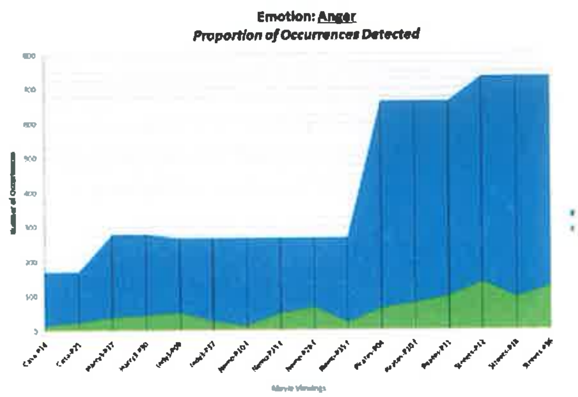
The diagrams in Figure 6.2(a) illustrate the proportion of annotated emotions actually detected using biometric measurements for each viewing. For each emotion category the number of occurrences of that emotion that were observed during the annotation of all six movies is plotted against the number of occurrences of the emotion detected using the biometric measurement for each viewing of those films. The blue areas in the diagrams below correspond to the observed or annotated occurrences, while the green areas illustrate the number of detected occurrences. The previously mentioned low average detection rate of 12% is plainly evident in the data illustrated by these diagrams.

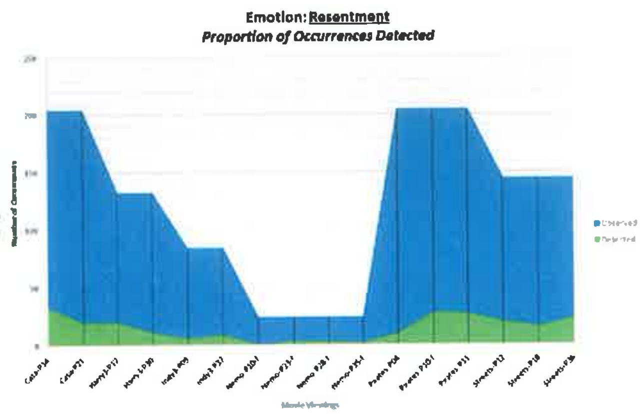


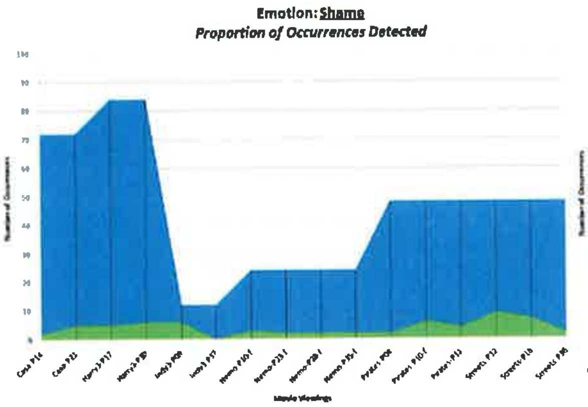


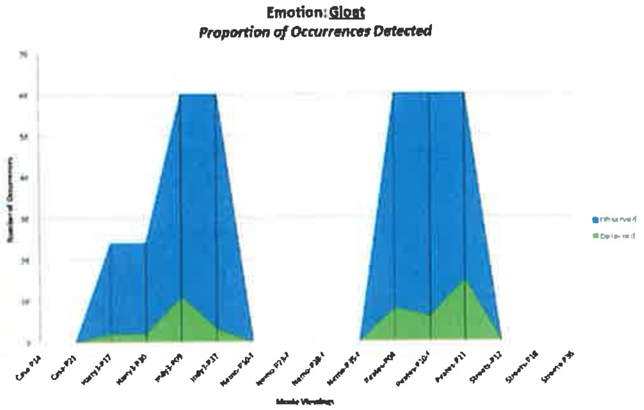


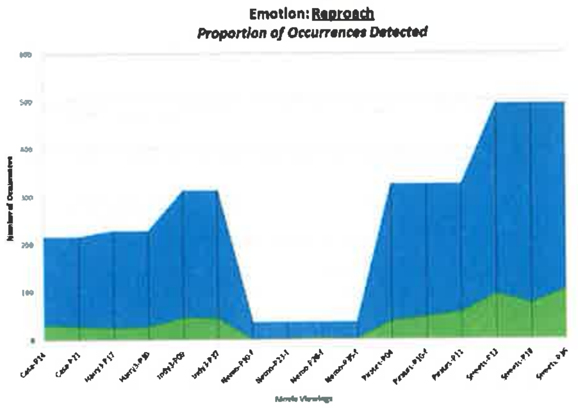


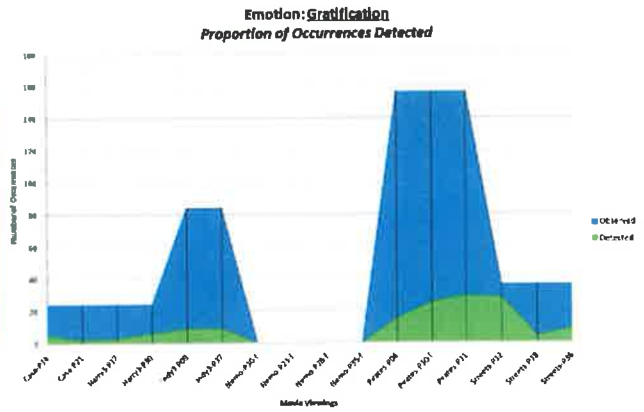


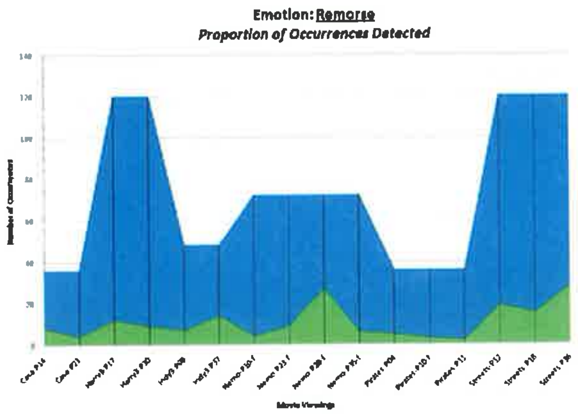


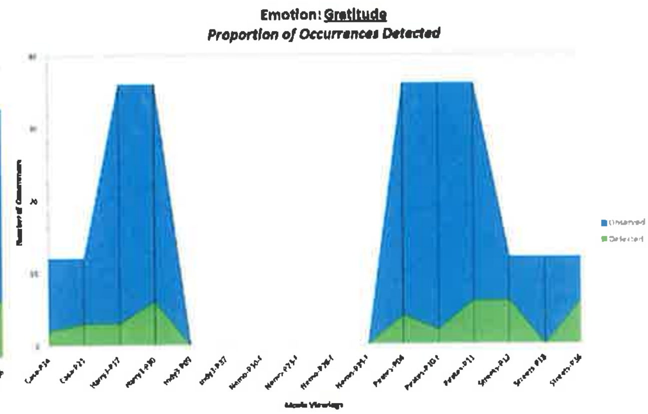

Emotion: Shame
Proportion of Occurrences Detected

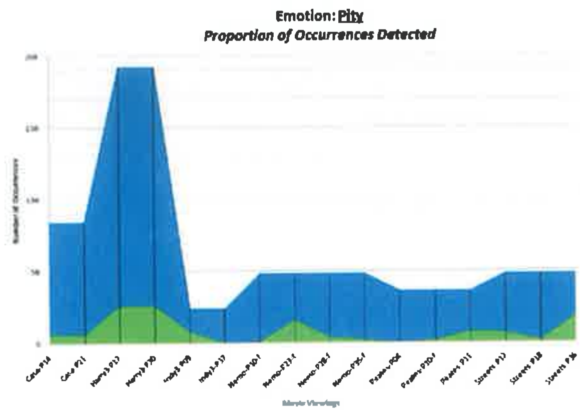

Emotion: Gloat
Proportion of Occurrences Detected

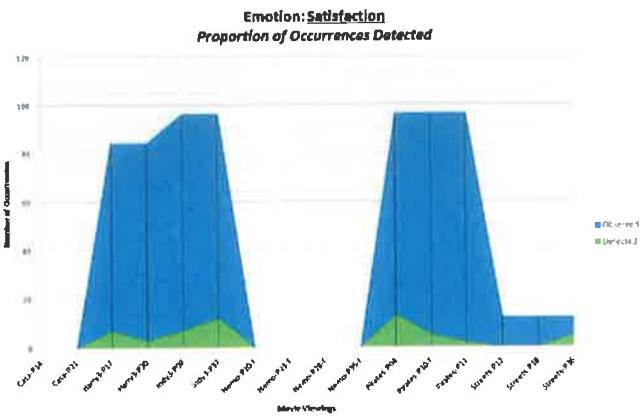

Emotion: Reproach
Proportion of Occurrences Detected

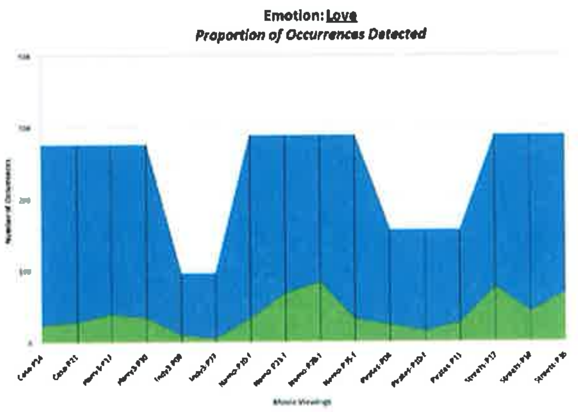

Emotion: Gratification
Proportion of Occurrences Detected

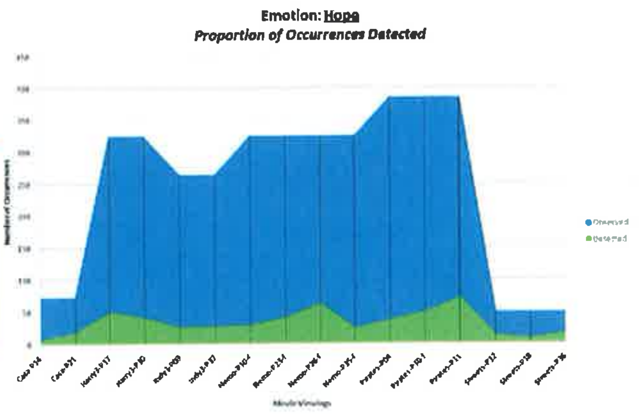

Emotion: Remorse
Proportion of Occurrences Detected

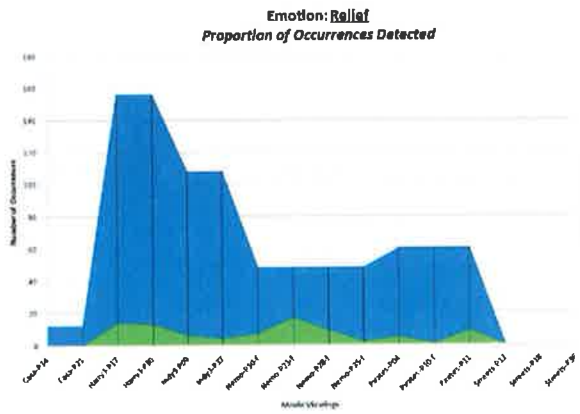

Emotion: Gratitude
Proportion of Occurrences Detected

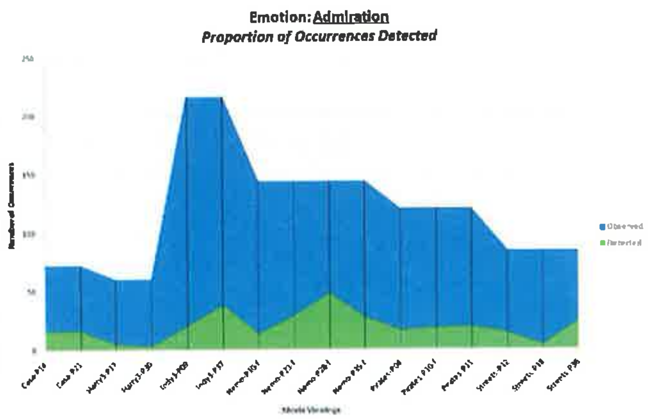

Emotion: Pity
Proportion of Occurrences Detected

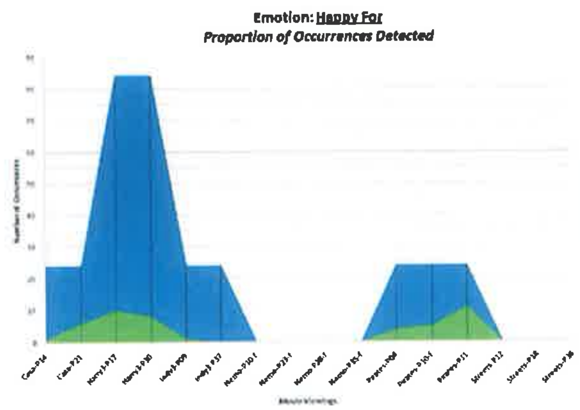

Emotion: Satisfaction
Proportion of Occurrences Detected

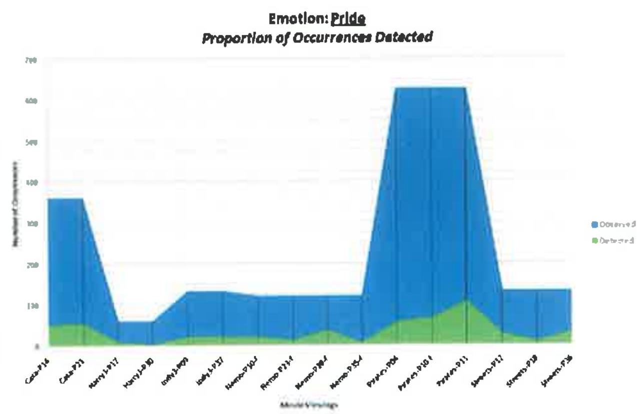

Emotion: Love
Proportion of Occurrences Detected


Emotion: Hope
Proportion of Occurrences Detected
Observed
Detected


Emotion: Relief
Proportion of Occurrences Detected


Emotion: Admiration
Proportion of Occurrences Detected
Observed
Detected


Emotion: Happy For
Proportion of Occurrences Detected


Emotion: Pride
Proportion of Occurrences Detected
Observed
Detected

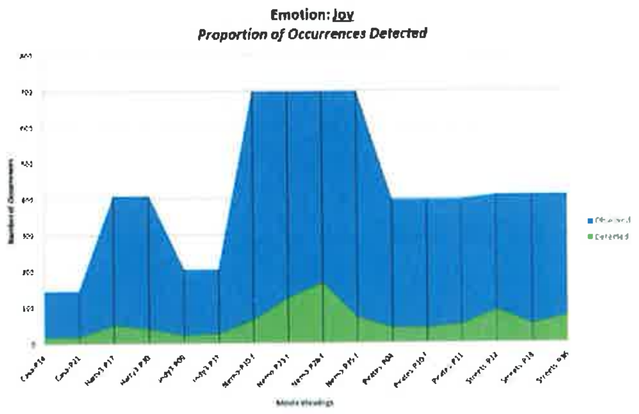


Figure 6.2(a): Proportion of Emotion Occurrences Detected Across All Movie Viewings

We can summarise these detection results by plotting the average number of observed emotion type occurrences for all viewings against the average number of detected emotion type occurrences for all viewings in Figure 6.2(b) below. From an examination of this diagram, it is apparent that there is a significant difference between the numbers of each emotion type annotated and the numbers of each emotion type detected using biometric measurements.

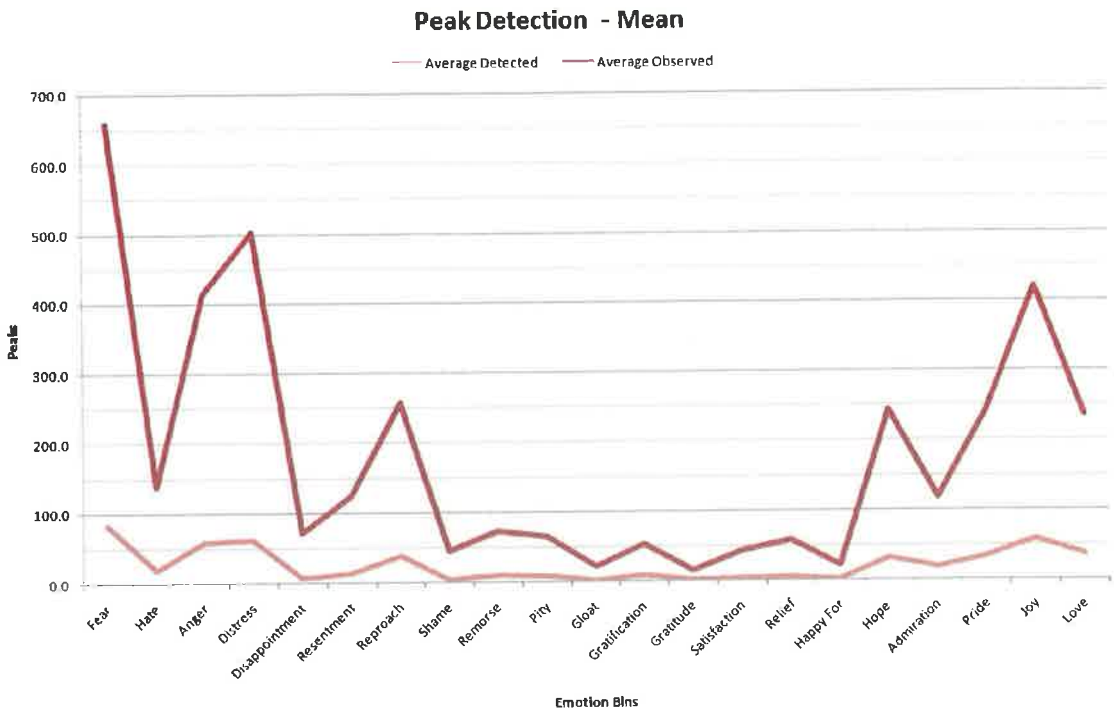


Figure 6.2(b): Overall Proportion of Emotion Occurrences Detected across all Movies

However, it is also apparent that there is some correlation between the two data series. In many instances, the number of detections can be seen to rise and fall approximately in line with the number of observed occurrences, however to a different extent.

It is this seeming correlation that we wish to investigate further. That is, if biometric techniques do not capture the sheer volume of emotion type occurrences, what is the nature of the relationship between what was detected and what was observed? Based on the diagram, Figure 6.2(b) above we can approximate that there is a correspondence between the two sets of data. In order to fully investigate any correlation between detections and observations we need to move away from the analysis of the volume of emotion type occurrences detected and observed, and analyse the results in terms of how the set of annotations relate to the set of detections.

6.2.2 Emotion Detection using Normalised Results

In order to investigate any correlation between the distribution of detections and the observations (or annotations) of emotion types, the two sets of data first had to be put into performance vectors with a value range of between 0 and 1, which is notated as [0, 1]. Normalisation is the process of transforming data from its value into a range such as [0, 1]. There are several ways to normalise a data series. To normalise a data series to one, the values have to be transformed into positive numbers, if necessary, and then divided by a value that is bigger than the nominator. Using this principle, any inequality can be used to normalise the data series. In this case, since the maximum and minimum value of both data series' are known, the transformation is as follows:

$$\delta_i = \frac{d_i - d^{min}}{d^{max} - d^{min}}$$

Equation 6.2(a): Normalisation

where the data series is in the range of $[d^{MIN}, d^{MAX}]$ and is not in the range of [0, 1], and δ is the vector of normalised values in the range of [0, 1] and i is the vector index. When $d = d^{MIN}$, then $\delta = 0$ and when $d = d^{MAX}$, then $\delta = 1$. As the values in both data series' are always ≥ 0, and the maximum value in each series is known, then we can set $d^{MIN} = 0$ and Equation 6.2(a) can be simplified into:

$$\delta_i = \frac{d_i}{d^{max}}$$

Equation 6.2(b): Simplified Normalisation

In this way, both data series of detections and observations are transformed into two vectors with values in the range [0, 1]. When these vectors are plotted against each other, in Figure 6.2(c) below, the effects of normalisation on the clarification of the detection and annotation results are startling, as 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 appears to almost exactly match that of the distribution of annotated emotion types.

These results would appear to support the hypothesis that emotional responses can be detected using biometric measurements. It is evident that while not all occurrences of an emotion in a movie can be detected by analysing the biometric results, with 12% of emotion occurrences being detected on average, any occurrences that *are* detected will parallel the same overall emotional pattern of the movie with 97% accuracy on average.

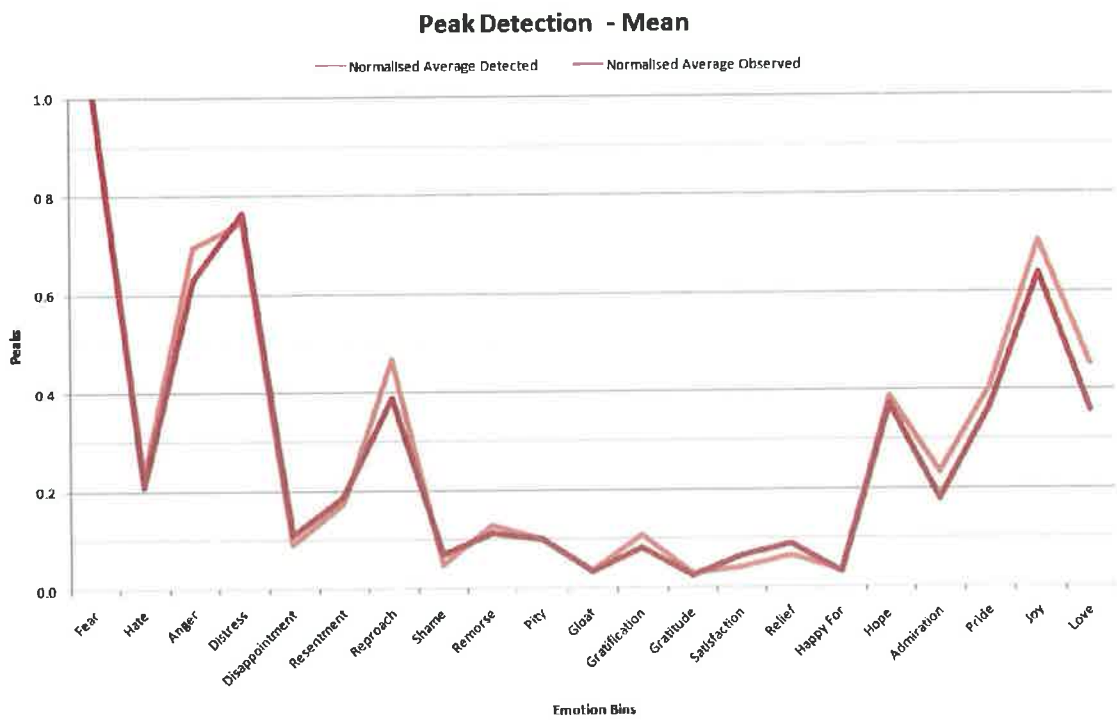


Figure 6.2(c): Emotion Detection vs. Emotion Annotation across all Movie Viewings

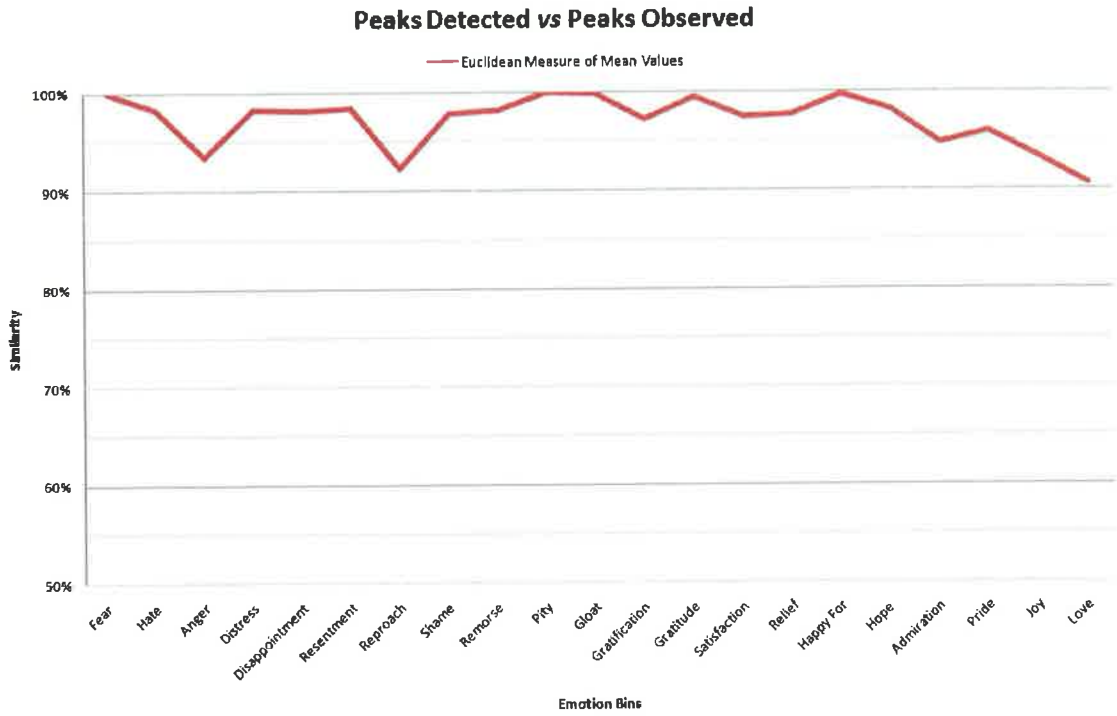


Figure 6.2(d): Correlation between the Automatic Detection and the Manual Observation of Emotions

The degree to which the detections parallel the observations for each emotion category is illustrated in Figure 6.2(d), where it can be seen that the degree of similarity between the two distributions never drops below 90%, and indeed only drops below 95% for four emotions, namely ANGER, REPROACH, JOY and LOVE.

Therefore it can be surmised that people do respond to emotions and that these responses can be detected using biometric measurements, just not on the scale of the annotated observations of each emotion type. What remains to be seen is what causes this discrepancy. Perhaps while people do respond physiologically to emotions, they don't respond at the same rate as was annotated. Or perhaps it is that the biometric techniques used in the CDV*Plex* experiment were not as accurate or comprehensive as was required. Perhaps other biometrics or more precise instrumentation could yield more success.

6.3 THE EFFECT OF MUSIC & EVENTS

In this section, we will investigate how the presence of music affected the emotional responses detected in film viewers and whether the different types of events which occur in a movie actually influence emotional reactions. In order to address our third and final hypothesis – which states that movie events and film music influence the emotional response of viewers – we will address such questions as; do people respond to emotions to a greater or lesser degree depending on the type of event that is occurring on-screen? What type of event most influences viewers' emotional responses? And does music actually impact viewers in a significant manner?

6.3.1 Music and Events

In previous chapters we have explored the elements that constitute a feature film. We began in Chapter 2 by considering movies as a source of emotional stimuli, and went on to explore the different devices used by filmmakers to establish and maintain an emotional connection with their audience. We discovered that movies owe a great deal of their ability to elicit emotional responses in viewers to the use of music. The psychological effects of music were examined and we explored the ways in which filmmakers exploit aspects of these effects for the purposes of plot exposition and audience engagement.

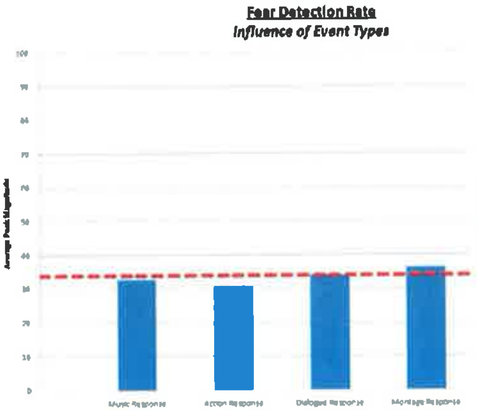
In Chapter 4, we considered the types of events that can occur in movies. An event is "something which progresses the story onward" [Lehane, 2006]; they are the pieces of a feature film that, when recalling a movie they have watched, a viewer will remember as a meaningful unit. We utilised the research conducted within our research group to identify the three fundamental types of event that occur in feature films, namely *action* events, *dialogue* events and *montage* events. *Dialogue* events are those that include people talking and form a major part of any movie. They are most often the means by which the film plot and background are conveyed to the viewer. *Action* or "exciting" events typically occur less often than dialogue events and include fights, car chases, battles etc. *Montage* events occur when spatially and/or temporally discontinuous shots are interwoven, usually with constant musical accompaniment. Emotional scenes, such as somebody crying or a romantic embrace, and montages are closely linked. Montages also include scenes of diegetic music such as a band playing a live song on-screen. Using the audiovisual analysis methods developed to automatically detect these three event types, we were able to identify the times at which music occurs on film soundtracks, and also ultimately to identify where in the films the different events occurred.

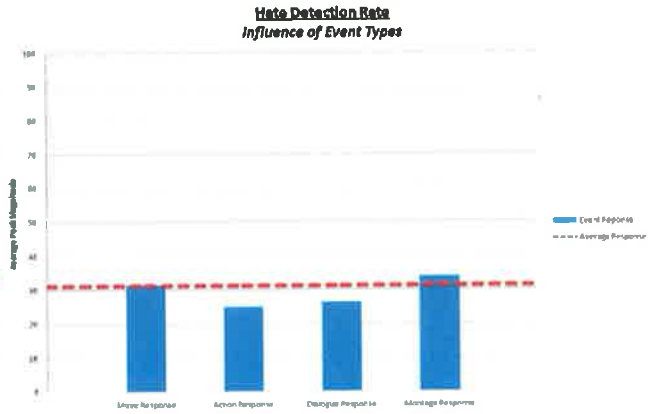
6.3.2 Music and Events' Influence on Emotional Responses

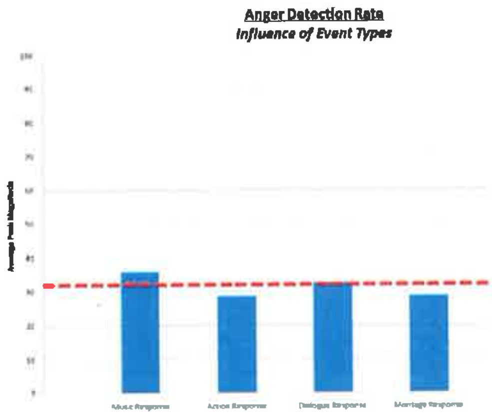
In this section, we will investigate whether movie viewers are emotionally influenced by the on-screen occurrences of different types of events or the presence of music. In order to measure the effect of music and events on how viewers react to emotion, we move away from biometrically detecting that a particular emotion had occurred, and instead measure the relative magnitude of the biometric peaks which were used to detect that emotion, where a magnitude of 100 is the largest peak that occurred for each biometric used and 0 is the smallest. In this way we are able to quantify whether people reacted to a greater, lesser or unchanged degree in the presence of music or certain events.

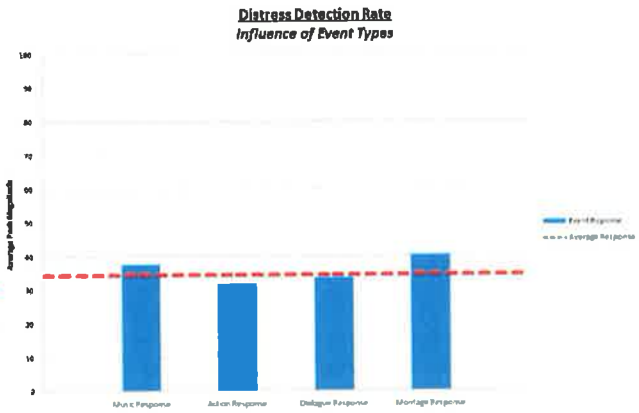
We will first consider the effect of music and events on the overall average biometric response magnitude per emotion. The following diagrams, shown below in Figure 6.3(a), detail the average size of the emotional responses, calculated for all occurrences of each type of event across all movies, and for the sections of the movies where music was detected on the soundtrack. The diagrams compare these averages against the overall average size of detected reactions for a particular emotion, and illustrate the effect of the occurrence of these events and the presence of music on the emotional responses and whether they increased, decreased or had no effect on the response magnitude.

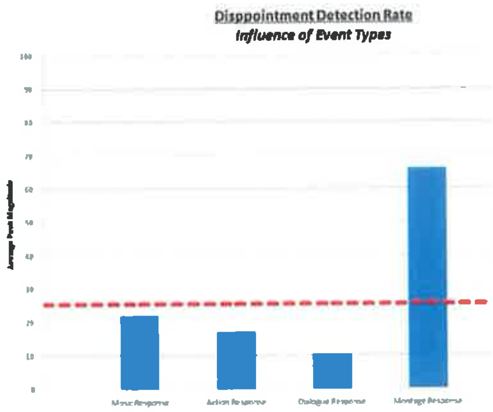
In the diagrams below, the red dotted line represents the average response for a particular emotion, regardless of the types of events which occurred or the presence of music across the six films. The blue columns represent the average response magnitude for music and each event type. When a column is observed to have crossed the average response line, that event's average response magnitude is greater than the overall average which signifies that the event has elicited stronger than average emotional reactions in viewers. For example, as illustrated below, the presence of music has elicited stronger responses for the emotions *ANGER* and *DISTRESS*, and montage events have elicited extremely strong responses for *DISAPPOINTMENT* and *GLOAT*.

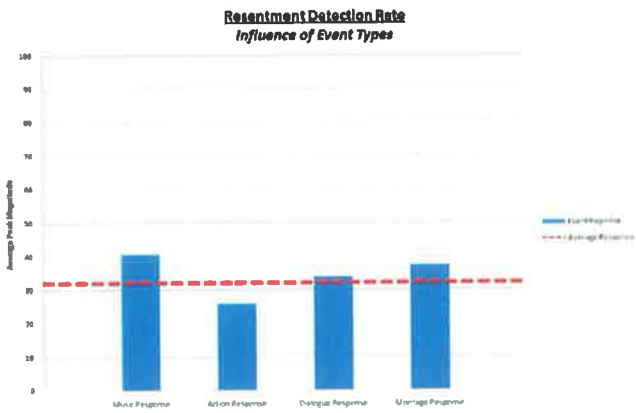


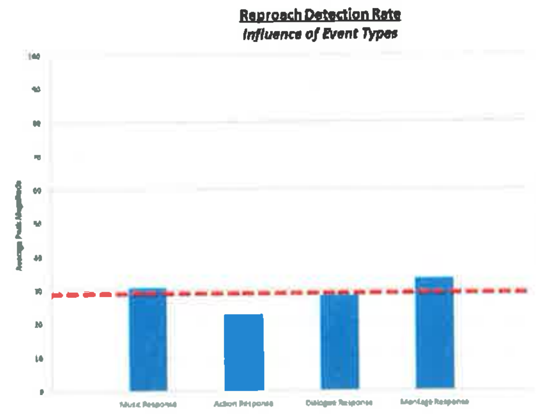


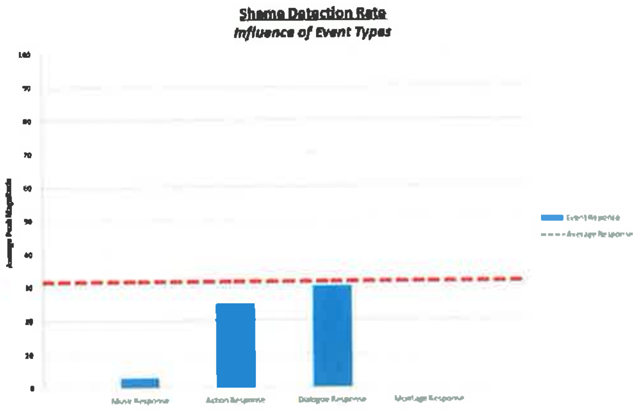


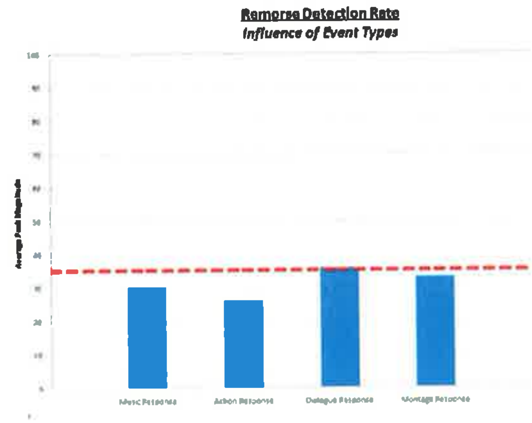


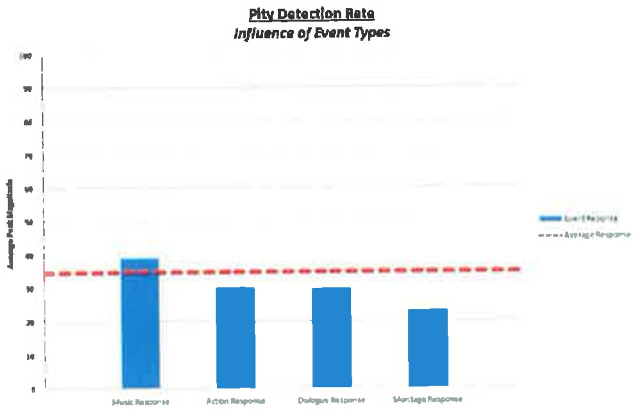


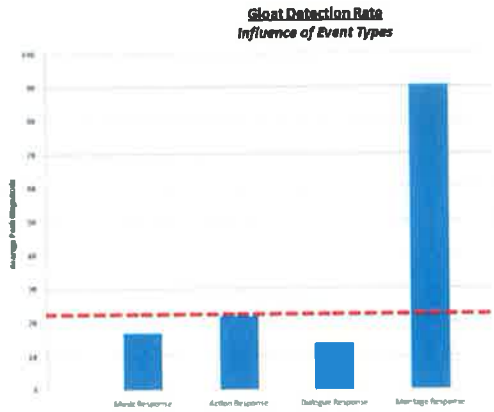


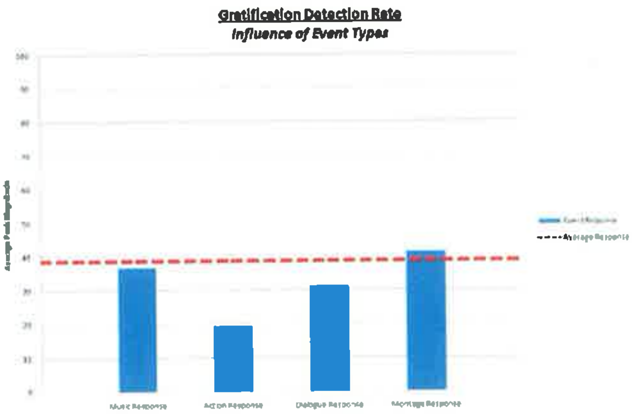

Reproach Detection Rate
Influence of Event Types
Music Response
Action Response
Dialogue Response
Montage Response

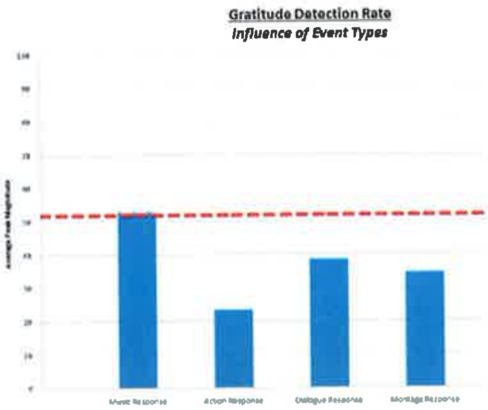

Shame Detection Rate
Influence of Event Types
Music Response
Action Response
Dialogue Response
Montage Response

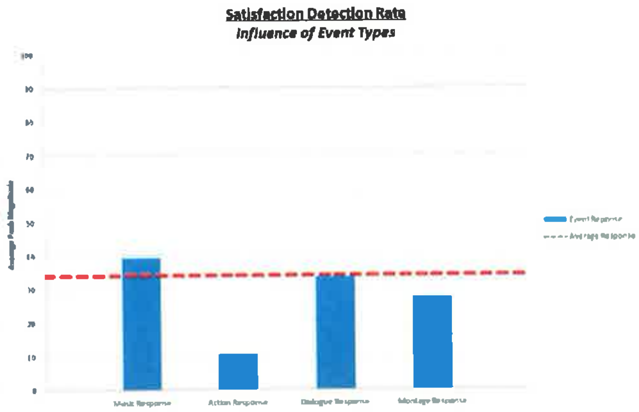

Remorse Detection Rate
Influence of Event Types
Music Response
Action Response
Dialogue Response
Montage Response

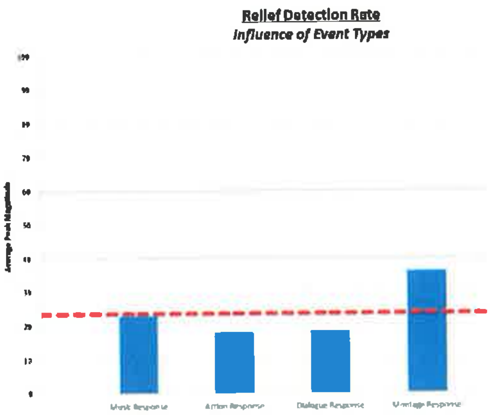

Pity Detection Rate
Influence of Event Types
Music Response
Action Response
Dialogue Response
Montage Response

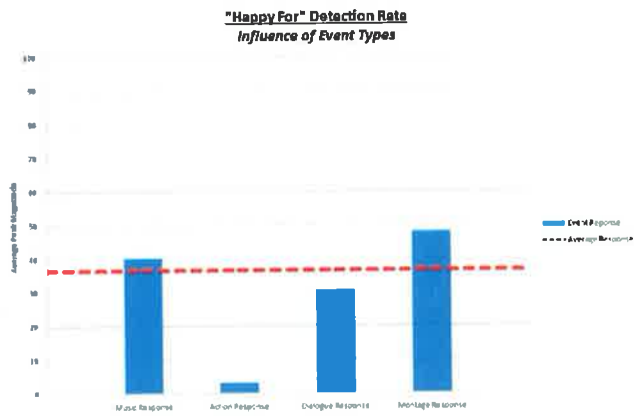

Gloat Detection Rate
Influence of Event Types

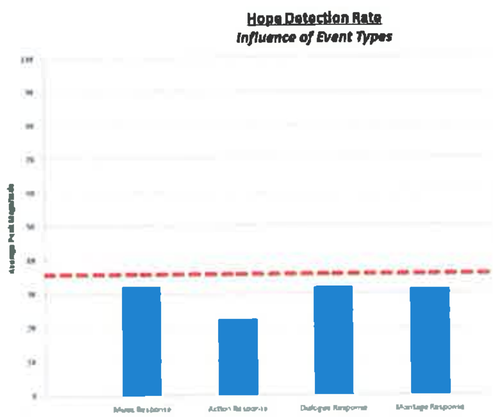

Gratification Detection Rate
Influence of Event Types

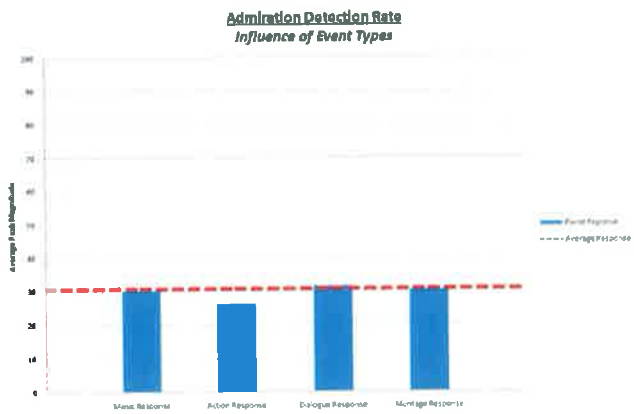

Gratitude Detection Rate
Influence of Event Types


Satisfaction Detection Rate
Influence of Event Types


Relief Detection Rate
Influence of Event Types


"Happy For" Detection Rate
Influence of Event Types


Hope Detection Rate
Influence of Event Types


Admiration Detection Rate
Influence of Event Types

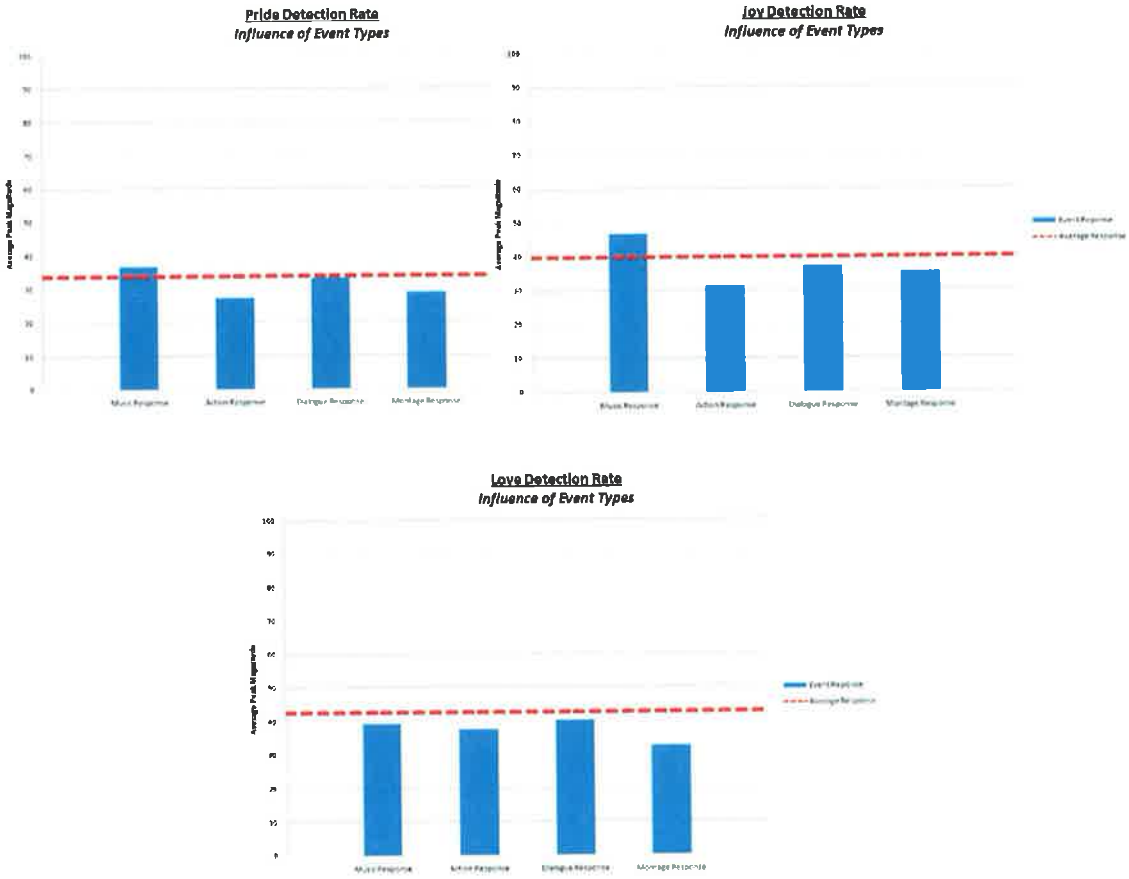


Figure 6.3(a): Variations of Average Responses per Emotion for Music & Events

However, in the cases where the blue columns fall below the average response then the music or event can be said to have dampened or diminished the response to an emotion. This is particularly evident for the emotion *SHAME*, where the presence of music and the occurrence of montages significantly dampened the emotional reactions of the viewers to that emotion, and similarly for action events on the emotion *HAPPY FOR*.

The diagrams shown in Figures 6.3(b), 6.3(c) and 6.3(d) below illustrate the proportion of the twenty-one emotion types for which the magnitude of viewers' responses increased, decreased and were not affected, respectively. For music and for each type of event, we calculated in what way the size of the viewers' responses to each type of emotion was influenced with respect to the average response magnitude. If the effect of the event or music was to increase or decrease the size of the reaction, for each emotion type, we computed whether the increase/decrease was less than 5% above/below the average (small), between 5% and 10% above/below the average (medium) or over 10% above/below the average (large). We then investigated whether the reactions for any emotion types remained completely unaffected by music and events; that is, where there were no changes in reaction magnitudes within 0% of the average, within 1% above and below the average and between 2% and 5% above and below the average. On studying these diagrams, the way in which emotion types are influenced by the presence of music or by the occurrence of certain events becomes plainly obvious. As we can see in Figures 6.3(d) and (e), for the majority of emotion types any influence music and events exert on viewers' reactions to these emotions is to decrease their magnitudes.

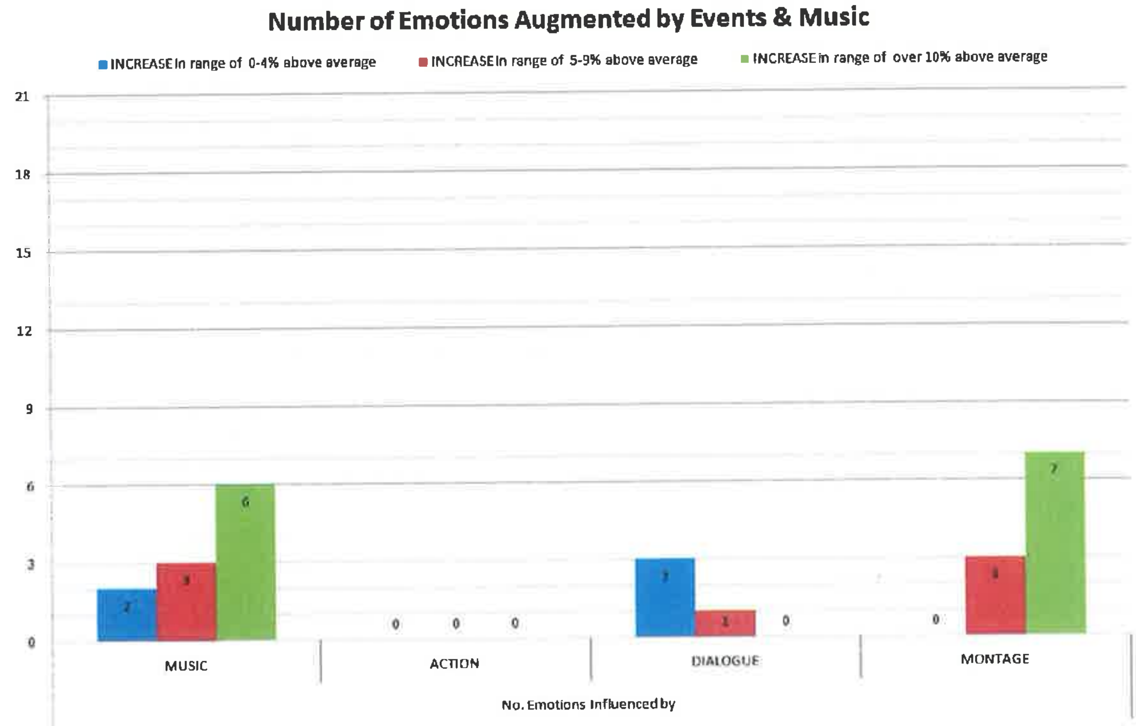


Figure 6.3(b): No. of Emotion Types with Increase in Responses for Music & Events

Action events in particular elicit reactions more than 10% lower than the average for nineteen types of emotion, that is 90% of all emotion types, and do not increase the strength of peoples' reactions for any emotion type. Dialogue events produce a small increase in response strength for three types of emotion and a medium increase for one emotion.

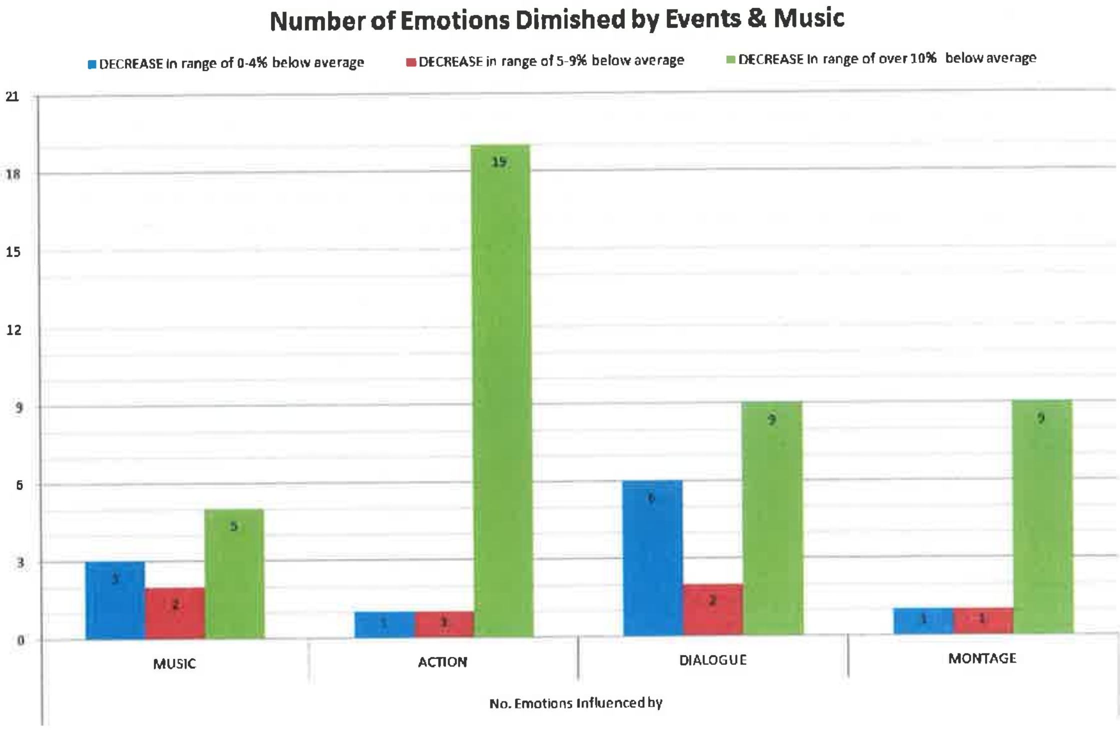


Figure 6.3(c): No. of Emotion Types with Decrease in Responses for Music & Events

Dialogue is more prone to decreasing the degree to which people react to emotions, demonstrating small decreases for six emotion types, medium decreases for two more emotions and larger decreases of more than 10% below the average for nine emotion types, which is 43% of all emotion types. It is with montages and music in general where we see peoples' responses increasing in strength for a larger percentage of emotion types.

Montages increase response strength in people for ten emotions (48% of all emotion types) and music for even more, with eleven emotions (52% of all emotion types). Montages seem to elicit slightly stronger reactions with medium increases for three emotion types, and large increases of more than 10% over the average for a record seven types of emotion, that is, one third of all emotion types. Music is not far behind this with large increases in six types of emotions, the same number of emotions with medium increases and smaller increases of 4% over the average or less for two emotion types. Music dampens responses in people for ten emotion types and montages for eleven emotion types but music has a greater proportion of smaller decreases than do montages. This is due to the fact that music elicits small decreases for three emotion types, medium decreases for two and larger decreases for just five emotion types, which are the lowest proportion of emotion types affected by decreases; whereas montages elicit small and medium decreases in just one emotion type respectively, but elicits decreases of more than 10% below average in nine types of emotion.

As we can see in Figure 6.3(d) below, there are no types of emotion which remain completely unaffected by music or events, though dialogue and music come close, to within 1% of the average for the emotion types of *FEAR* and *HATE* respectively as can be observed in the diagrams of Figure 6.3(a). This means that the small way in which dialogue affects how viewers react to *FEAR* comes to within 1% of the average reaction to *FEAR*, and similarly for how music affects viewers' responses to *HATE*. When the range is then further extended to as much as 5% either side of the average response size, we see that one or two more emotions fall into the "unaffected" category, in that any influence music and events have on viewers reactions is within 5% or less of the average response.

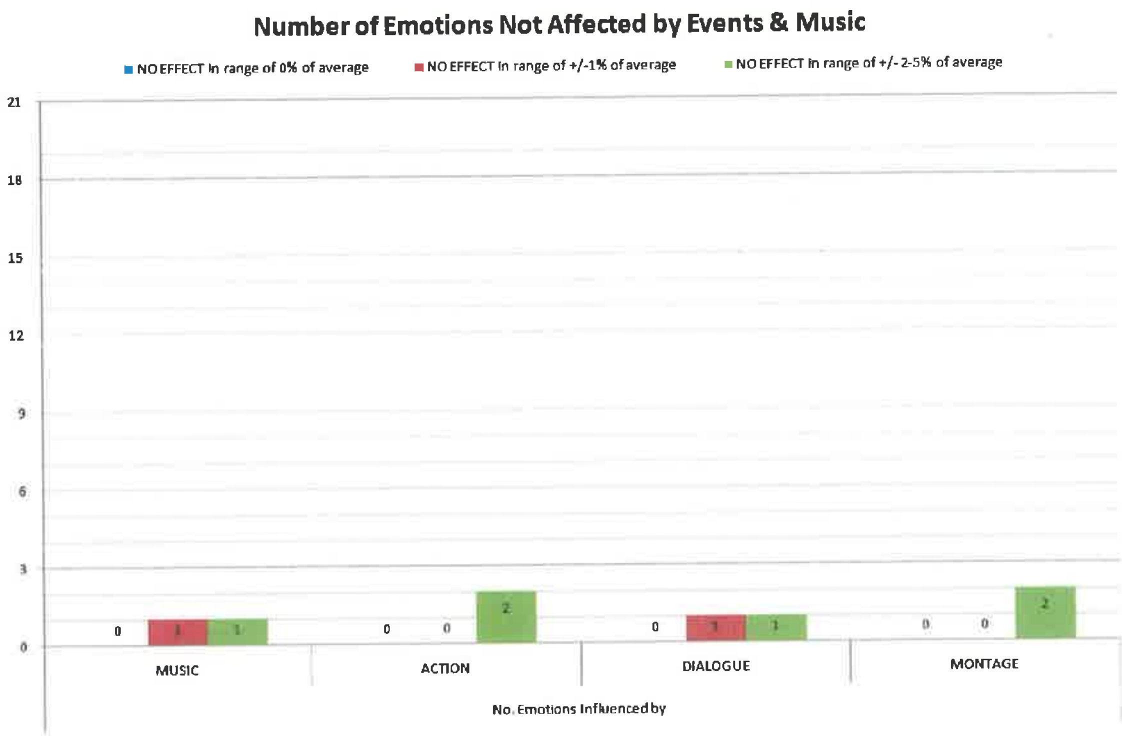


Figure 6.3(d): No. of Emotion Types Unaffected by Music & Events

These results address some of the questions relating to our third hypothesis, which states that movie events and film music influence the emotional response of viewers. Does music actually impact viewers in a significant manner? We have seen that music elicits stronger responses for the largest percentage of emotion types, and has the least dampening effect for the smallest proportion of emotion types in which viewer responses were diminished. What type of event most influences viewers' emotional responses? Montage events, with their high degree of music content, elicit the strongest reactions for the greatest proportion of emotion types, while action events exhibit the strongest dampening effect on viewers' reactions to over 90% of emotion types. Do people respond to emotions to a greater or lesser degree depending on the type of event that is occurring on-screen? It is this question on which we will focus in the next section.

6.3.3 Degree of Emotional Influence of Music and Events

In the previous section we demonstrated that music and events do influence viewers' emotional reactions while watching movies. We investigated the types of emotions which are affected by the presence of music or the occurrence of events, and which of these exerted a boosting or dampening effect on the magnitude of viewers' responses to these emotion types. We observed that the presence of music significantly impacts viewers for the majority of emotion types, and discovered that montage and action events exert the most significant influence on the largest proportion of emotion types, though the effects are the quite opposite. Montage events tend to augment the magnitude of viewers' reactions and action events significantly diminish the magnitude of viewers' reactions.

The diagram shown below in Figure 6.3(e) illustrates the degree to which people respond to emotion, with respect to the presence of music or type of event that is occurring on-screen. Each data series represents the distance (or percentage effect) of the average response magnitude for music and events from the overall average response magnitude, for each emotion type. The overall average response is therefore portrayed as a straight line at 0%. In this way, the degree of influence of music and each event is clearly depicted, ranging between over 300% above average and nearly 100% below average, and the scale of the effect they exert on the size and strength of viewers' reactions to different types of emotion is evident.

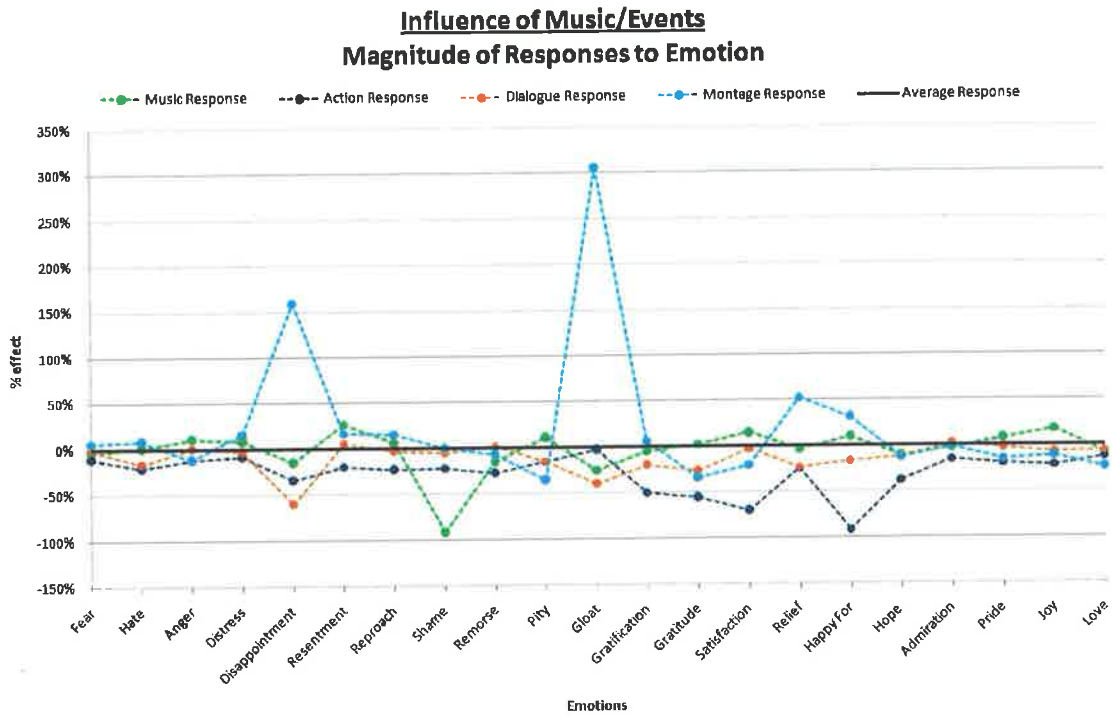


Figure 6.3(e): Influence of Music & Events on Emotional Responses in Viewers

We previously determined that, for more types of emotions than any type of event, music increased how strongly people react to those emotions. We also determined that montages also increased how strongly people react to emotions, though for less emotion types than music, but generally eliciting stronger reactions. This conclusion is also borne out with these analyses, where we can actually see the size of the effects which music and montage have on the magnitude of peoples' reactions. Music elicits increases of up to 27% above the average magnitude, while montages far exceed this with some very large increases of up to just over 300% above average. However if we temporarily exclude the two significant outlying increases of 300% for GLOAT and 159% for DISAPPOINTMENT, the next largest increase for montages in 52%, which is more in line with the other results.

In order to address our final question, relating to our third hypothesis, which asks whether people respond to emotions to a greater or lesser degree depending on the type of event that is occurring on-screen. The results tell us that yes they do, and to quite a significant degree. While action events have a very definite dampening effect on people's reactions to emotions, dialogue events also have a general dampening effect, displaying an augmenting effect for just four types of emotion. Montages however have a much less dampening effect and for a large proportion of emotion types, nearly 50% of all emotion types, actually elicit stronger reactions from people – for one emotion a reaction magnitude that is three times the average magnitude. What we observe is that music and the one type of event in which music is a dominant factor, elicit much stronger reactions from people for more types of emotions than do other types of events.

Thus our findings support our third and final hypothesis, that movie events and film music do influence the emotional response of viewers. We find that the presence of music, and events which are music rich generally make viewers react more strongly to emotions.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

This chapter will briefly review the work carried out and discussed in this thesis chapter by chapter, before presenting an overview of our findings and conclusions. We will also outline some ideas for possible future work.

7.1 THESIS SUMMARY

The thesis begins by introducing, in Chapter 1, the concept of detecting emotion in film, as well as providing some background information on the area of Information Retrieval and it's applications to multimedia content. We discussed video information retrieval and the applications of the work described in this thesis in this area. We proposed the hypotheses addressed in this thesis; firstly, that emotional responses experienced by people watching movies can be detected using biometric measurements; secondly, that individual viewers experience similar reactions to emotional stimuli; and finally, that movie music and filmic events in a movie influence viewers' emotional responses.

Chapter 2 then went on to describe how music is used in feature films, and to what effect. We discussed the psychological impact that music can have on the human mind, and how people perceive music. The effect music has on our emotions was also considered, appreciation of which is fundamental to understanding how and why music is used in feature films. This chapter then conclude with an overview of film scoring.

Chapter 3 described the CDV*Plex*, an experiment we conducted in order to compile a ground-truth of human subject responses to stimuli in feature films, by monitoring and recording the physiological reactions of people as they viewed a feature film in a controlled cinema-like environment. This chapter introduced the set of biometric measurements we recorded for each participant and described the cinematic environment and experimental procedure that was followed. The results of the biometric analysis and the peak detection processes were then presented and explicated.

Chapter 4 described research conducted within our research group and pertaining to the work carried out in this thesis, in which an approach for automatically indexing the content of feature films was developed. The research described here involved using an examination of the common filmmaking techniques in order to derive a "grammar" for feature films. This derived grammar was combined with automatic audio-visual analysis of a feature film, based on the extraction of low-level descriptive features, in order to facilitate the automatic detection of high-level semantic movie events such as dialogue, action and montage events.

Chapter 5 described the process by which a selection of feature films were manually annotated for the presence of emotional content. We explained our choice of the set of emotion types to mark up each feature film, and then proceeded to describe the movie annotation process in detail. We then discussed each of the six selected movies, including a synopsis of each storyline and then finally, we examined the results of the annotation.

Chapter 6 presented and examined the results of the experiments carried out and discussed in this work and described our findings. We evaluated the data extracted from the biometric measurements recorded in the CDVPlex experiment and investigated the findings with respect to the three fundamental hypotheses put forward in this work. A comprehensive summary of these findings

7.2 SUMMARY OF FINDINGS

We have analysed the results and data gathered in our experiments in order to address the three hypotheses which form the basis for this work. Our first hypothesis states that emotional responses experienced by people watching movies can be detected using biometric measurements. For each of the twenty-one categories of emotions used in this work, we have found that 12% of occurrences of these emotions are being detected on average using biometric measurements. However, while not all occurrences of each emotion in a film can be detected by analysing biometric results, any occurrences that *are* detected will parallel the same overall emotional pattern of the movie with 97% accuracy. That is, we have found the distribution of detected emotional responses to be almost identical to the distribution of the actual observed emotion occurrences. We can therefore conclude there is evidence to support our first hypothesis, as according to our analysis emotional responses in viewers can be detected biometrically.

Our second hypothesis states that individual viewers experience similar reactions to emotional stimuli. We began by comparing the reactions of people who had watched the same film and found that, in all cases but one, the degree of similarity between viewers of the same film was over 90% - in one case as high as 96%. The one exception to this was a similarity measure that was still very high at 89%. Therefore we concluded that when people view the same film they respond in a similar manner. We also observed that the rates of peak detection for each film all fall within a 5% range, and so we calculated the overall similarity of all viewings for all films. We found that for nearly 40% of emotion types there was over 95% similarity of responses, and an overall average similarity of responses for all emotion types of 92%. We conclude that these results support the hypothesis that viewers experience similar reactions to the same emotions, whether they are all viewing the same material or not.

Finally, our third hypothesis states that movie music and filmic events in a movie influence viewers' emotional responses. For music and each event type, dialogue, action and montages, we determined whether a boosting or dampening effect was exerted on the magnitude of viewers' responses. We observed that the presence of music significantly impacts viewers for the majority of emotion types, in that it elicits stronger responses of up to 27% above average for the largest percentage of emotion types. Music also has the least dampening effect for the smallest proportion of emotion types in which viewer responses were diminished. We then discovered that montage events, the one type of event in which music is a dominant factor, elicit much stronger reactions – some as large as three times the average – from people for more types of emotions than do other types of events. We also determined that action events have a very definite dampening effect on people's reactions to 90% of emotion types and that dialogue events also have a general dampening effect. We therefore conclude these results support the hypothesis that the presence of music and the occurrence of events influence viewers' emotional responses as we have established that music has a significant influence in increasing the strength of viewers' reactions to emotions and that this contributes to the strong augmentation effect of montage events, while action events exert the most significant dampening influence on the largest proportion of emotion types.

We have therefore established that people experience similar and measureable physiological responses to emotional stimuli in feature films, and that these responses are influenced by the presence of music and the occurrence of different types of events in the film, and that we can detect these responses through analysis of their biometric measurements.

7.3 FUTURE WORK

Research into emotion detection and analysis of movies presented in this thesis demonstrates scope for research to continue into emotion-based analysis and detection for the video domain and in this section we suggest future work and possible extensions arising out of this work.

In terms of the experimentation carried out and discussed in this work, and the results obtained from this, it would be interesting to expand our analysis to encompass a greater number of movies. In this work, we constrained ourselves to six films of varied genre in order to develop a set of processes by which to analyse the information captured in association with a movie, such as our emotion annotation process, biometric collection and peak detection processes, and the automatic event and music detection processes discussed in Chapter 4. In applying these processes to a wider variety of films across many genres, it would be interesting to discover the commonalities across films of the same genre and era in terms of emotion detection.

Similarly, a worthwhile further experiment would be to expand the number of viewers for each film for which biometric data was collected. In this work we used two to four people per film and discovered that generally people react in a similar manner. Increasing this number to twenty, ten or even five people per film was just not possible given the time constraints and the availability of equipment. In the future it would be interesting to conduct larger scale CDV*Plex*-type experiment in which more viewers watched each movie, and each movie was given several, rather than one, separate showings.

The range of biometric measurements we used, though comprehensive, was not by any means exhaustive. It would be interesting to test whether alternative biometrics, such as electromyogram (EMG), electrocardiogram (EKG), blood pressure and respiration, would improve on our emotion detection rate of 12%. EMG measures the electrical activity of muscles. When muscles are active, they produce an electrical current that is usually proportional to the level of muscle activity. EKG measures the electrical activity of the heartbeat. Respiration measures the breathing rate. However these biometrics require complex and expensive instrumentation in order to measure and record responses, and certainly neither EMG nor EKG instrumentation would be wearable nor unobtrusive. Respiration and perhaps blood pressure would be a more practical next step.

In terms of practical applications of the findings in this work, the benefits of using emotions as a means of semantically indexing fictional video have been expounded in Chapter 1. The advantages to video information retrieval would be manifold, improving video indexing and summarisation techniques, so that they would behave in a more intuitive and meaningful manner for the user. It would enable movies to be represented by a breakdown of the emotions it induces in its viewers, thus users could search and browse movies, and perhaps other fictional video, based on their emotional content. In a real world setting, this has obvious applications in movie recommender systems, in that it would allow users to retrieve films based on their desired type of film or movie experience.

APPENDIX A

MUSIC THEORY

A.1 PITCH

Musical pitch refers to the perception of the frequency of a note. It is used to describe the concepts of melody, harmony, tonality, range, and tuning. The pitch that attains the greatest stability in a musical passage is called the tonal centre.

A.1.1 Timbre

Timbre is the quality of a sound, as determined by its frequency spectrum, and includes tone colour and articulation. It varies between voices and types of musical instruments. A tonal, or pure, sound is produced when sound waves are periodic (pulse at regular intervals). The human voice has a tonal quality when singing. A noisy sound is made up of a mix of frequencies which produce very complex waveforms. Musical instruments generally emit several frequencies, known as harmonics, which have regular waveform patterns and these equate to the timbre of the instrument, and can range from the simple harmonics of a flute or triangle to the complex ones of the violin or timpani.

A.1.2 Scales, Keys and Modes

A scale is an ordered collection of notes capable of maintaining a consistent set of relationships. In conventional Western music, a scale consists of eight notes whose pitches progress from a root note and the root's first octave, separated by intervals of tones and semitones (the smallest interval commonly used in Western music). Humans accustomed to tonal music like to categorise musical notes into a twelve-tone equal-tuned scale (i.e. chromatic scale), where each note is just a semitone higher than the one preceding it. When a note strays from this perfect pitch, it is regarded as being "out-of-tune", as it is wavering between two semitones. Most conventional Western music is based on the diatonic system, where the pitch intervals between consecutive tones in the diatonic scale are 2, 2, 1, 2, 2, 2 & 1 semitone(s).

A musical "key" then refers to the tonal centre of a musical piece. The keys used in the Western tradition are two modes of the diatonic scale; a major key that uses the first note of the diatonic scale as the root and a minor key that uses the sixth note of the diatonic scale as the root. Other modes are also possible but most modern popular music and European music of the eighteenth- and nineteenth-centuries uses the diatonic modes. The early music of Greek antiquity referred to scales in the context of scalar modes rather than modern "keys", which were named after ancient Greek cities that supposedly preferred a given mode in times past. The modern conception of modal scales describes a system where each mode is the usual diatonic scale, but with a different root note. Three of the modes are based on the major scale, while four of them are based on the minor scale. Much folk music is best analysed in terms of modes, for example the Ionian, Dorian, Aeolian and Mixolydian modes occur in decreasing order of frequency in Irish traditional music, and the Phrygian mode is an important part of the Spanish flamenco sound.

A piece may modulate or change key at some point. Modulation is sometimes implemented by simply starting in the new key with no preparation - this kind of key change is common in various kinds of popular music, when a sudden change to a key a semitone or whole tone higher is a quite frequently heard device at the end of a song. In classical music, however, a "smoother" kind of key change is more usual.

A.1.3 Harmony

Harmony is the relationship between two or more simultaneous pitches, that is, the relation of notes to notes and chords to chords as they are played simultaneously, where a chord is two or more different notes or pitches sounding simultaneously, or nearly simultaneously, over a period of time. Whereas melodic intervals are those that are linear and occur in sequence, harmonic intervals are sounded at the same time. Whether or not a harmony is pleasing is a matter of cultural or even personal taste, as there are consonant and dissonant harmonies, both of which are pleasing to the ears of some and not others ["The Four Elements of Music", 2005].

A.2 RHYTHM

Rhythm is, by its simplest definition, musical time. It is the organization of the durational aspects of music. Regular pulsations in the music are called the beat. Tempo is the musical term that indicates the overall pace of an arrangement. Tempos range from *grave*, meaning solemn and extremely slow, to *allegro*, meaning fast and cheerful.

A.2.1 Metre

Metre is the pattern of accentuated beats, which is repetitive and predictable - it gives order to time. For example, the three beat measures of a waltz indicate triple metre (3/4) and the four beat measures of a march indicate common (quadruple) metre (4/4). Without metre, music sounds like a Gregorian chant. When the melody falls on notes that occur between beats it said to be "syncopated" time.

A.2.2 Phrasing

Phrasing is the rhythm of organic movement; it arises naturally from the need to breathe while singing or playing a wind instrument. Without it, music would sound mechanical and become boring.

A.3 MELODY

> *"It is melody that enables us to distinguish one work from another. It is melody that human beings are innately able to reproduce by singing, humming and whistling. It is melody that makes music memorable: we are likely to recall a tune long after we have forgotten its text"* [Selfridge-Field, 1998]

Together with rhythm, harmony and timbre, melody is one of the main dimensions of musical description. It carries implicit information regarding harmony and rhythm and has great importance in music perception and understanding. Several instruments may play interleaved notes, creating the illusion of a single melody (fusion). Conversely, a single instrument may play alternate notes from very distinct frequency ranges, thus appearing to be playing two different melodies at once (fission), as is often heard in Bach's compositions.

A.4 GENRE

A musical genre is a category of musical composition characterised by a particular form, style or content. Some genres, such as Hungarian folk music, are geographically defined while others, like Baroque, Classical and Romantic music, are largely defined by a time period. Still others, such as Gregorian chant, are defined by quite precise technical requirements. Music can be divided into genres in many different ways and these classifications are often arbitrary and can be controversial, with closely related styles often overlapping. Soifer's "Directory of Musical Styles" [1997] lists musical styles in chronological order, with their periods and musical characteristics, as well as some representative composers, performers and compositions.

A.5 NARRATIVE

In the Classical and Romantic periods of music of the 19th century, storytelling was inherent in purely instrumental pieces, with distinctive architectures relating to a beginning, middle and end. The *sonata* is a typical example of this, where two thematic ideas are presented in the opening section (exposition), then in the middle section (development) these are brought to tension through key changes with transforming or opposing themes, and finally in the last section (recapitulation) there is a return to the original themes and key, bringing the piece to a conclusion. Thus, just as music composition has its own structure, when accompanying a story it can provide a formal unity by employing repetition, variation and counterpoint, thus supporting the narrative. An example of how music alone can convey a story are to be found in Prokofiev's "Peter and the Wolf" where each character in the folk tale is illustrated by specific instrumentation and melody, and the story is very effectively played out by the interaction of these purely instrumental musical elements.

APPENDIX B

CDV*PLEX* FILMOGRAPY

Film	Year	Genre
Alien	1979	Action/Adventure
Indiana Jones and the Raiders of the Lost Ark	1981	Action/Adventure
Indiana Jones and the Temple of Doom	1984	Action/Adventure
Indiana Jones and the Last Crusade	***1989***	***Action/Adventure***
Clockers	1995	Action/Adventure
Crouching Tiger, Hidden Dragon	2000	Action/Adventure
Pirates of the Caribbean: The Curse of the Black Pearl	***2003***	***Action/Adventure***
Who Framed Roger Rabbit?	1988	Animation/Comedy
Toy Story	1995	Animation/Comedy
Shrek	2001	Animation/Comedy
Finding Nemo	***2003***	***Animation/Comedy***
Happiness	1998	Comedy
Disco Pigs	2001	Comedy
Roger and Me	1989	Documentary
Lawrence of Arabia	1962	Drama
The Godfather I	1972	Drama
The Godfather II	1974	Drama
Veronica Guerin	1979	Drama
Chariots of Fire	1981	Drama
The Godfather III	1990	Drama
Schindler's List	1993	Drama
Michael Collins	1996	Drama
Nora	2000	Drama
Apocalypse Now	2003	Drama
Das Boot	1981	Foreign
Il Postino	1994	Foreign
The Son's Room	2001	Foreign

The Changeling	1980	Horror/Suspense
The Shining	1980	Horror/Suspense
Casablanca	***1942***	***Romance***
Doctor Zhivago	1965	Romance
The English Patient	1996	Romance
Titanic	1997	Romance
Chocolat	2000	Romance
2001: A Space Odyssey	1968	Sci-Fi/Fantasy
Star Wars IV A New Hope	1977	Sci-Fi/Fantasy
Star Wars V The Empire Strikes Back	1980	Sci-Fi/Fantasy
E.T. the Extraterrestrial	1982	Sci-Fi/Fantasy
Start Wars VI Return of the Jedi	1983	Sci-Fi/Fantasy
Harry Potter and the Philosopher's Stone	2001	Sci-Fi/Fantasy
The Lord of the Rings: The Fellowship of the Ring	2001	Sci-Fi/Fantasy
Harry Potter and the Chamber of Secrets	2002	Sci-Fi/Fantasy
The Lord of the Rings: The Two Towers	2002	Sci-Fi/Fantasy
The Lord of the Rings: The Return of the King	2003	Sci-Fi/Fantasy
Harry Potter and the Prisoner of Azkaban	***2004***	***Sci-Fi/Fantasy***
Mean Streets	***1973***	***Thriller/Mystery***
Taxi Driver	1976	Thriller/Mystery
Leon	1994	Thriller/Mystery
Pulp Fiction	1994	Thriller/Mystery
Jackie Brown	1997	Thriller/Mystery
Minority Report	2002	Thriller/Mystery

APPENDIX C

CDV*PLEX* DOCUMENTATION

DCU

DUBLIN CITY UNIVERSITY

Plain Language Statement of the study "Analysing Movies"

I. Introduction to the Research Study

The Centre for Digitial Video Processing is conducting a research study to correlate human subject responses to stimuli in movies, with observed biometric responses. The study is led by Prof. Alan Smeaton and colleagues.

II. Details of what involvement in the Research Study will require

Participants in the research study are invited to view a movies, of their choice, in a cinematic environment (big screen, darkened room, surround sound, etc.) in groups of four. Prior to the movie showing they wear an armband which measures and logs galvanic skin response, heat flux, and accelerometer movement, as well as a sports heart rate monitor in the form of a strap around their chest. Data from these devices, gathered during the movie is then uploaded for analysis. The benefit you get from taking part in this study is that you get to see a movie, for free!

III. Confidentiality of data

Data provided by you will be rendered anonymous by assigning a master code number to each participant and the mapping between participant codes and participant names will be kept by the principal investigator only. This data will be subject to legal limitations and will be destroyed after a 5 year period.

IV. Nature of the Research Study

Participation in the research study is voluntary and participants may withdraw from the Research Study at any point.

If participants have concerns about this study and wish to contact an independent person, please contact:

The Secretary, Dublin City University Research Ethics Committee, c/o Office of the Vice-President for Research, Dublin City University, Dublin 9. Tel 01-7008000

DCU

DUBLIN CITY UNIVERSITY
Informed Consent Form

I. Research Study Title: Analysing Movies; Centre for Digital Video Processing

II. Purpose of the Research: The research to be carried out is to correlate human reaction to highlights in movies, with automatically-derived highlights.

III. Confirmation of particular requirements as highlighted in the Plain Language Statement

Participant – please complete the following (Circle Yes or No for each question)

Have you read or had read to you the Plain Language Statement	Yes/No
Do you understand the information provided?	Yes/No
Have you had an opportunity to ask questions and discuss this study?	Yes/No
Have you received satisfactory answers to all your questions?	Yes/No

IV. Confirmation that involvement in the Research Study is voluntary

Participants may withdraw from the Research Study at any point.

V. P{lease note that the confidentiality of data, including that confidentiality of information provided is subject to legal limitations

VI. Signature:

I have read and understood the information in this form. My questions and concerns have been answered by the researchers, and I have a copy of this consent form. Therefore, I consent to take part in this research project

Participants Signature: ______________________________

Name in Block Capitals: ______________________________

Witness: ______________________________

Date: ______________________________

CDVPlex EXPERIMENT

Date:	*Movie:*	*Device number:*

BEFORE VIEWING THE FILM

Personal Details

Your full name: ______________________________

Your gender:

❑ Male ❑ Female

I am a:

❑ smoker

❑ non-smoker

Your height: __________

Your weight: __________

I am:

❑ Dominantly right-handed

❑ Dominantly left-handed

❑ Both

Your age:

❑ Less than 20
❑ 20 – 24
❑ 25 – 29
❑ 30 – 34
❑ 35 – 39
❑ 40 – 44
❑ 45 – 49
❑ 50 – 54
❑ 55 – 59
❑ 60 – 64
❑ More than 64

About your film preferences....

Have you seen this film before? ❑ Yes ❑ No

If yes, when?

❑ Less than 1 month ago

❑ Less than 6 months ago

❑ About 1 year ago

❑ About 2-5 years ago

❑ About 6-10 years ago

❑ More than 10 years ago

How often do you go to the cinema?

❑ Seldom ❑ 1-2 times a year ❑ Once a month ❑ 2-3 times a month ❑ Once a week ❑ More than once a week

What genre do you like? (Score between 1 and 5: 1= not at all; 5= very much)

Action ______	Animation ______	Comedy ______
Drama ______	Sci-Fi ______	Thriller ______
Foreign ______	Romantic ______	Documentary ______
Horror ______	Other ______ (specify: ______________________________)	

Sit back, relax and enjoy the film !!

AFTER VIEWING THE FILM

We hope you enjoyed the viewing...

Which aspect of the film did you enjoy the most?
(Score between 1 and 5: 1= not at all; 5= very much)

Storyline _____ Actors/Actresses _____ Music ____
Sound effects _____ Other ____ (specify: ____________)
Overall score _____

Write down the highlights of the film (and how it made you feel):

1.

2.

3.

4.

5.

Thank you very much for your participation.

APPENDIX D

DETAILS OF ANALYSED FEATURE FILMS

<u>Finding Nemo</u> *(2003)*

Genre: Comedy/Animation

Directors: Andrew Staunton, Lee Unkrich

Composer: Thomas Newman

In 2003, "Finding Nemo" became the world's highest grossing animated film of all time and the ninth highest grossing film ever made, and received an Academy Award, or "Oscar", [Academy of Motion Picture Arts & Sciences, 2006] for Best Animated Film and was nominated for three more Oscars for Best Original Screenplay, Best Original Score and Best Sound Editing [Internet Movie Database (IMDb), 2006].

The film follows the comedic and eventful journeys of two clown fish, Marlin and his young son Nemo who live in the Great Barrier Reef. The tale begins with Marlin enduring a predator attack which kills his wife and their rather large family of newly laid eggs, of which the damaged egg containing Nemo was the only survivor. The action then skips to a few years later, to an excited Nemo's first day of school. When an overprotective Marlin frets about Nemo's ability to look after himself due to his damaged fin, Nemo rebels and swims out into the open ocean on a dare. To Marlin's horror, Nemo is captured by a scuba diver and taken away on a boat, leaving Marlin with one clue – an address written on a pair of goggles, dropped by the diver.

A frantic Marlin embarks on a dangerous trek across an ocean and finds himself the unlikely hero of an epic journey to rescue his son, who has ended up in a dentist's fish tank overlooking Sydney Harbour. Marlin teams up with the eternally optimistic but dangerously forgetful fish called Dory and together they overcome many trials and tribulations, involving sharks with acute fish-addictions, an explosive World War II submarine, toxic jellyfish, some friendly, thrill-seeking turtles and a rather mysterious whale. Meanwhile, Nemo and his tank-mates scheme on how to get out of the tank before Nemo has to face his fate as a gift intended for the dentist's little niece, who unfortunately has a bad track record in keeping her pets alive. Following a number of failed escape attempts, a local pelican called Nigel, befriended by Nemo, hears of Marlin's story which has spread all around the ocean and a determined Nemo makes one final effort by playing dead. Unfortunately, having been rescued from a flock of marauding seagulls by Nigel, Marlin and Dory arrive in time to see this charade and in the confusion are convinced Nemo really is dead and leave, grief-stricken. Having being successfully flushed down a drain, Nemo finally reaches the ocean where he bumps into a distressed Dory. Together they go after Marlin, who's gone off to be alone. Despite a heart-felt reunion, their troubles aren't over as they're suddenly caught in a huge fishing-net with a school of terrified fish. However Nemo saves the day and leads them all to freedom and a happy ending, as Marlin and Nemo return home.

Pirates of the Caribbean: The Curse of the Black Pearl ***(2003)***

Genre: Action/Adventure

Director: Gore Verbinski

Composer: Klaus Badelt and Various

Take a film full of pirates, a damsel in distress and a romantic hero, and throw in a cursed ghost ship, and you have the swashbuckling tale set in 17^{th} century West Indies that is the "Pirates of the Caribbean". However, this is an old style film with a modern twist – the damsel doesn't need all that much rescuing, the romantic hero is a lowly apprentice blacksmith and the pirates really aren't all that bad…those that remain curse-free anyway. A huge blockbuster in 2003, "Pirates of the Caribbean" was nominated for five Oscars including Best Make-Up, Best Special Effects, Best Sound, Best Sound Editing and Best Actor for Johnny Depp in his outrageously comic portrayal of the swaggering rogue, Captain Jack Sparrow [IMDb, 2006].

The film opens with Governor Swann and his young daughter Elizabeth travelling from England to Port Royal in the Caribbean under the command of a young Captain Norrington, when they come across a burning ship with an unconscious young boy as the sole survivor. He is Will Turner and when Elizabeth discovers a gold pirate medallion on him she assumes him to be a pirate boy and takes the piece to protect him from being hanged for his supposed association. The action then skips eight years into the future and Elizabeth is a beautiful young woman with the gold medallion still hidden in her possession, Captain Norrington is being promoted to Commodore and Will, now an apprentice blacksmith, is very much but very inappropriately in love with the lovely "Miss Swann". At the cliff-top ceremony, just as the Commodore is proposing to her, Elizabeth faints in the heat and falls hundreds of feet into the sea whereupon the medallion she's still wearing reacts curiously with the water to send out a pulse of energy - ominously, the wind changes direction. She is rescued by the rather unconventional pirate Captain Jack Sparrow who, having arrived in port aboard a sinking ship, was at the nearby dock bamboozling two rather hapless soldiers and eyeing up the *Interceptor*, pride of the British fleet. Jack is immediately arrested for being a pirate, to be hanged in the morning, but his daring escape leads him to Will's forge where he is recaptured, after a prolonged swordfight with Will. However, overnight the town is raided by the dreaded ghost ship the *Black Pearl* – previously Jack's ship before he was mutinied – and Elizabeth is kidnapped by the pirates, who have a special interest in her gold medallion.

Will is determined to rescue her, and breaks Jack out of gaol whereupon, with some ingenious misdirection, they steal the *Interceptor* and sail on to the pirate stronghold of Tortuga to pick up a crew. Along the way, Jack reveals to Will the devastating news that Will is the only child of "Bootstrap" Bill Turner, a pirate who sailed with Jack on the *Black Pearl*. Meanwhile, Elizabeth learns the true meaning of the curse of the *Black Pearl* as the moonlight reveals the crew to be walking, rotting corpses who cannot be killed. Her gold piece and her blood, as the supposed Turner child, are the keys to lifting their curse on the mystical *Isla de Muerta*. However it doesn't work, and Will, the rescued Elizabeth and the medallion escape in the *Interceptor*, hotly pursued by the enraged pirates who now hold Jack prisoner. The pirate Captain Barbossa captures and sinks the Interceptor, taking captive her crew and Will, his true identity now known. Jack and Elizabeth are marooned on a desert island but are shortly rescued by her father and the Commodore after she lights a huge signal fire using a smuggler's cache of rum. By accepting the commodore's marriage proposal, she persuades him to stop Barbossa and rescue Will. An epic battle ensues in which undead pirates battle British Redcoats, Jack battles Barbossa and Elizabeth first rescues Jack's crew and then Will. Barbossa is tricked and the pirates are beaten.

With his crew and his reacquired ship fled, Jack is to be sent to the gallows despite a protesting Elizabeth. However, Will intervenes and after declaring his love to Elizabeth, finally, attempts to rescue Jack. They make a last stand and confront Norrington, when Elizabeth steps in and, declaring her love for Will, pleads for mercy and lenience. The film ends with Jack diving from the cliff, just in time to be picked up by the *Black Pearl* and escape to freedom. The Commodore graciously yields to Elizabeth's decision and, with her father's blessing, she and Will embrace at last.

<u>Indiana Jones and the Last Crusade</u> ***(1989)***

Genre: Action/Adventure

Director: Steven Spielberg

Composer: John Williams

The third instalment of the perennial favourite Indiana Jones trilogy sees actor Harrison Ford return to his signature role as the archaeologist/adventurer, in the top grossing film of 1989. Embarking on another epic quest, Indiana Jones must find his estranged father who's gone missing while searching for the Holy Grail. The film won an Oscar for Best Sound Effects Editing, and was also nominated for Best Original Music Score and Best Sound [IMDb, 2006].

The film opens in 1912, with Henry "Indiana" Jones Jr. as a young Boy Scout trying to save the historical Cross of Coronado from grave robbers he discovers while on a field trip, escaping the looters by way of a moving circus train where he first picks up his trademark whip, his fear of snakes, and a scar on his chin, as well as his iconic fedora hat.

The story then moves on to 1938, and a now adult Indy, having tracked down the looters, is on their ship off the Portuguese coast. He finally retrieves the Cross and donates it to the museum run by Marcus Brody, a family friend.

Back at the university, having narrowly escaped being mobbed by his own students, Indy is summoned to meet the wealthy Walter Donovan, who informs him that his father, Henry Jones Sr., has vanished while searching for the missing half of a clue to the location of the Holy Grail. Indy and Marcus travel to Venice to meet his father's partner, Dr. Elsa Schneider, in order to retrace his father's footsteps. Inside the library where Henry Sr. was last seen, Indy finds the entrance to the underground catacombs for which his father was searching, by using his father's "Grail Diary" containing all he'd learned about the Grail over a lifetime, and which he'd posted to Indy just days before. He discovers the tomb of Sir Richard, a knight of the First Crusade, which holds information needed to find the Grail. However, Indy and Elsa are pursued by members of "The Brotherhood of the Cruciform Sword"; a secret cult whose mission is to protect the Grail at all costs. In the ensuing motorboat chase Indy kills all except the leader, Kazim whom he convinces that he is simply looking for his father, not the Grail. Kazim tells him that his father is being held in Castle Brunwald near the border between Austria and Germany.

Marcus goes on to Turkey to help Indy's friend Sallah prepare for the Grail quest, while Indy goes to find his father, but they are betrayed by Elsa and Donovan, who are working in league with the Nazis. His father's kidnapping had been a ploy to get Indy to solve the mystery of the Grail for them. Indy and his father escape and go to Berlin, to retrieve the diary which Elsa had stolen. Meanwhile, the Nazis capture Marcus in Iskenderun, Turkey, and thus learn of the starting point for the quest.

Travelling from Iskenderun, Donovan and Elsa take Marcus with them but are tracked down by Indy, Henry Sr. and Sallah who plan on rescuing him. Kazim and his men attempt to stop the Nazi caravan, but outgunned and outnumbered, they are killed. In the confusion, Henry Sr. attempts to rescue Marcus from inside a tank, but is himself captured, leading Indy to jump onto the tank and rescue both of them after a prolonged fight with Donovan's aide, the evil Nazi Colonel Vogel, who is finally killed when the tank drives off a cliff.

The quest reaches its climax at the nearby Canyon of the Crescent Moon, the site of the temple housing the Grail. Donovan captures Indy in the temple and shoots Henry Sr., forcing Indy to retrieve the Grail to use its healing powers to save his father's life. Indy, guided by the diary, circumvents the deadly booby-traps, reaching a room where a knight of the First Crusade, kept alive by the power of the Grail, has hidden it amongst many false cups. Donovan and Elsa follow Indy to the room, where she identifies a golden, bejewelled cup as the Grail and Donovan impatiently drinks from it to become, as he thinks, immortal. It turns out to be her way of ridding herself of a rival for the Grail as Donovan dies in a very grotesque manner, aging rapidly into dust.

Indy picks out the true Grail, a plain cup, and proves it by drinking from it without harm. The knight warns him not to let the Grail leave the temple before Indy fills the Grail with water rushes and back to heal his father. However, Elsa then tries to leave with the Grail and the temple starts to collapse. As the ground splits open, she falls into a crevasse and instead of letting Indy pull her to safety, she greedily reaches for the Grail which has fallen just out of reach, and falls to her death. Indy then loses his footing and finds himself in the same situation. With Henry Sr. holding him in place Indy also tries to get the Grail, but his father tells him to let it go and, unlike Elsa, he reluctantly obeys. The Grail is lost forever as they escape the crumbling temple but they are safe and reunited, and the film ends with Indy, his father, Sallah and Marcus riding off into the sunset.

<u>Harry Potter and the Prisoner of Azkaban</u> ***(2004)***

Genre: Science-Fiction/Fantasy

Director: Alfonso Cuarón

Composer: John Williams

The series of "Harry Potter" books by J. K. Rowling, and the associated films, describe a world where magic abounds in an ancient and hidden subculture of witches, wizards and various magical and fantastic creatures, of which the average person knows nothing despite living alongside it. Harry achieved fame at the age of one when Lord Voldemort, the most feared and evil wizard to have ever come to power, attacked his home. Voldemort murdered Harry's parents but mysteriously fails to kill Harry as Voldemort's curse rebounded, reducing him to a weak, disembodied spirit and leaving Harry with a distinctive scar on his forehead. Raised an orphan by his vile, neglectful and non-magical aunt and uncle, Harry didn't discover his wizard heritage until his eleventh birthday when he was admitted to Hogwarts School of Witchcraft and Wizardry, a British boarding school for children with magical power. The third film in the hugely popular Harry Potter series introduces a darker tone than the previous films, mirroring the progression into darker territory in the source book, as Harry and his friends enter their third year at Hogwarts. "Harry Potter and the Prisoner of Azkaban" was nominated for two Academy Awards, Best Original Score and Best Visual Effects [IMDb, 2006].

The film begins with Harry at home with his odious relatives, the Dursleys, during his school holidays, when his Uncle Vernon's formidable and extremely antagonistic sister Marge comes to visit. It all ends in disaster as Harry, under great provocation, finally loses control of his anger and therefore his magic when Marge callously insults his tragically deceased parents. The accidental result of this sees Marge inflating like a balloon and floating away into the sky. A shocked and angry Harry storms out of the house with his trunk of possessions in tow, and having accidentally hailed the magical "Knight Bus", a purple triple-decker, he travels to the magical Leaky Cauldron pub in London. Here, much to his bemusement, the surprisingly relieved Minster for Magic awaits him to tell him that the "Marge situation" has been handled and that there were no further repercussions for Harry, even though he had broken the strictly enforced ban on under-age wizards using magic outside of school. It seems that Harry is in great danger from Sirius Black, a dangerous murderer and zealous follower of Lord Voldemort, who has escaped from the terrible wizard prison of Azkaban after twelve years of incarceration. What's worse is that it is later revealed that Black is also Harry's godfather, and was once his father's best friend, before he betrayed Harry's parents to Voldemort.

Harry and his best friends Ron Weasley and Hermione Granger start their year at Hogwarts in a traumatic manner, having been attacked on the school train by Dementors who are ostensibly looking for Black. These horrific creatures are the guards of Azkaban whose very presence is enough to induce the reliving of the worst memories of a person's life. In Harry's case, their presence causes him to collapse as he remembers his mother screaming in her final moments. The new "Defence Against the Dark Arts" teacher, Prof. Remus Lupin, drives them off and they finally reach Hogwarts in safety, where Headmaster

Dumbledore announces that Dementors will be patrolling Hogwarts perimeter until Black was caught.

Upon his return to school, however, Harry is relatively unconcerned with Black, as Dumbledore is widely regarded as the most powerful wizard of the age and Hogwarts is renowned for its safety. However, Harry struggles to thwart the Dementors who seemed to take an unhealthy interest in him. Lupin, another old friend of Harry's parents, teaches Harry to conjure the very advanced Patronus Charm as protection from his tormentors.

A confrontation between Harry and Black is inevitable but, when it finally happens, the young wizard is shocked by revelations of Black's innocence and the deviousness of Ron's pet rat Scabbers, who it turns out is really the wizard Pettigrew in disguise. Pettigrew, an old friend of Lupin's, Black's and the Potters' long presumed dead, was the traitor who truly betrayed Harry's parents to their deaths. Sirius looks set to be exonerated and Harry regains a Godfather who cares for him. But Lupin hides a dark secret of his own - he is a werewolf and as they step out into the light of the full moon holding Pettigrew captive, Lupin starts to change into his out-of-control beast form. As they scatter, Pettigrew escapes, and Harry and Sirius are attacked by Dementors but are saved by a mysterious wizard who incants a very powerful *Patronus Charm.* As Pettigrew was the only proof of Sirius' innocence, Sirius is arrested and sentenced to have his soul sucked out by a Dementor. However Dumbledore intervenes and Harry and Hermione are transported a few hours back in time to effect a number of rescues, including Sirius'. Harry is revealed to be the mysterious powerful wizard and effectively saves his past self and Sirius from the Dementors. Black escapes to freedom thanks to Harry and Hermione, but Lupin must resign as the fact that he is a werewolf has become public knowledge. A depressed Harry receives an anonymous gift of the world's best and latest model racing broom and the film ends with Harry joyfully soaring above Hogwarts.

<u>Casablanca</u> ***(1942)***

Genre: Romance/Drama

Director: Michael Curtiz

Composer: Max Steiner

The classic and much-loved film "Casablanca" is a romantic story set during World War II in the French-Vichy controlled Moroccan city of the same name. It stars the Hollywood greats Humphrey Bogart and Ingrid Bergman in one of the most memorable on-screen romances of all time. It won an Oscar for Best Film as well as two more for Best Director and Best Screenplay. It was also nominated for five more, including Best Actor for Bogart as Rick and Best Score for the legendary Max Steiner [IMDb, 2006]. When Germany invaded France, many French citizens and resident foreigners tried to flee to America through Lisbon. Once the route to Lisbon was blocked by the Nazis, the only escape route left was through North Africa, specifically Casablanca. Because of the high cost of travelling and difficulty obtaining the correct documents, many refugees were trapped in Casablanca, desperately trying to find a way out.

As the story begins, it is revealed that two Nazi couriers have been murdered for their travel permits, documents that could easily get a person out of Casablanca. Coincidentally, a shady character enters Rick's American Café in Casablanca and asks the owner Rick Blaine to keep safe a couple of recently acquired and very valuable travel permits. Rick reluctantly agrees. The corrupt French chief of police, Louis Renault, informs Rick that the Resistance leader Victor Laszlo, a Czech, has escaped from a concentration camp and is now in Casablanca. Legally, the Nazis can't do anything against Laszlo on unoccupied French soil but they want to stop him from getting to America. It is suspected that the stolen travel permits will be used to help Victor Laszlo and his wife get out of Casablanca. Rick is determined to remain neutral and denies knowing anything. The Nazis arrive soon afterwards to arrest the man who gave Rick the permits and he is killed.

Victor Laszlo and his Norwegian wife, Ilsa Lund, arrive at the Café and as they are seated she and Sam, the piano-player and Rick's long-time friend, are shocked to recognise each other. We learn that a few years previously in Paris, just before the Nazi invasion, Rick and Ilsa were madly in love. They were supposed to leave France together for Casablanca, but Ilsa left him waiting at the train station with a note, not explaining herself but swearing that she really did love him. Rick travelled on to Casablanca alone, heartbroken and emotionally hardened, thinking he'd never see her again. He is still deeply hurt by what had happened and refuses to talk to Ilsa, driving her away with harsh and cold words.

Victor and Ilsa desperately need to get out of Casablanca, but to do that, they need the permits that it is now rumoured Rick has in his possession. When Rick denies Laszlo's request, Ilsa returns to Rick to beg him but he refuses. Ilsa breaks down and confesses that she still loves him. They reconcile, and she explains how she had married Victor before she and Rick had met in Paris, but she thought he'd been killed by the Nazis. Then, just before Rick and she were to leave for Casablanca, she learned that Victor was still alive, and she had had to go him. Now, however, all she wants to do is stay with Rick and let her husband go on to America without her, but she's in turmoil.

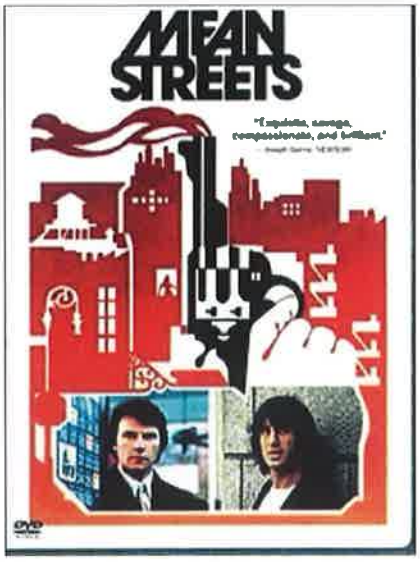
Rick begins planning what appears to be an escape for Ilsa and himself from Casablanca. He goes to Renault and plans to have Victor arrested, leaving Rick and Ilsa free to use the travel permits to leave Casablanca. However, at the last moment, Rick double-crosses Renault and uses him to guarantee Victor and Ilsa's safe passage to the plane to Lisbon. Ilsa thinks at first that she's staying behind with Rick, but Rick convinces her that the right thing to do is for her to go with Victor saying, with that famous line, that she would regret staying "Maybe not today. Maybe not tomorrow, but soon and for the rest of your life". She's stricken, but leaves - both know that they will always love each other. Rick plans to carry on the fight against the Nazis in his own way, as a mercenary. Renault also has a surprising change of heart about the Nazis and the final scene sees Rick and Renault walk off together into the dense fog covering the airfield to go join the Free French fighters.

Mean Streets ***(1973)***

Genre: Thriller/Mystery

Director: Martin Scorsese

Composer: Various



"Mean Streets" is an early Scorsese film starring young up-and-coming actors Harvey Keitel and Robert DeNiro. Made on a budget of just US$500,000 it was one of Scorsese's first successes, he having also written the story and screenplay. The focus of the film is the conflict within the central character, an outsider whose views on life are a little different from those around him. It has a violent, gritty atmosphere and the entire film takes place in the Little Italy area of New York in the 1970s, an ugly and criminal-infested neighbourhood filled with violent hoodlums, loan sharks and gangsters.

Virtually plot-less, *Mean Streets* meanders through a series of tableaux which give a powerfully realistic impression of life in Little Italy. The film centres on the struggles of four such residents of Little Italy, all men in their mid-20s who seem to share genuine affection and friendship. Big friendly Tony runs the local bar and strip-club; Michael is a loan shark who also likes to scam and rob naive teenagers from Brooklyn; Johnny Boy is an angry, crazy, irresponsible hooligan with a penchant for blowing things up, and borrowing money from loan sharks whom he never intends to repay. The main protagonist, Charlie, is the well-dressed nephew of the local mafia boss Giovanni whom he works for by collecting debts, but who wants nothing more than to run his own restaurant.

Charlie is torn between the life of the streets and the life his uncle can give him. He is also deeply religious, with a deep admiration for St. Francis of Assisi, and constantly inflicts self-prescribed physical penitence for his sins on the streets by holding his fingers over a flame. Charlie is hampered by his feeling of responsibility towards his childish yet destructive friend Johnny Boy. He is also having a hidden love affair with Johnny Boy's cousin, Teresa, who is an epileptic and is ostracised because of her condition - especially by Charlie's uncle. Giovanni, doesn't want Charlie to keep company with either Johnny Boy or his pretty cousin Teresa - Johnny Boy because he seems to be so unstable, and Teresa because of her epilepsy. However, Charlie is determined to "save" Johnny Boy even though he seems to be a lost cause. Constantly late for his weekly loan payments, violent and disrespectful to everyone, he gets Charlie deeper and deeper into trouble. Life continues in its normal fashion, with its routine violent fights at pool-halls & card games, going to the movies, heavy drinking and the celebration of a traditional religious feast. Until Johnny Boy pushes his luck to the limits and the film comes to a tragic and violent conclusion.

Throughout the film Charlie is constantly running interference between his friend Michael and Johnny Boy, smoothing the loan shark's ruffled feathers. But as Johnny Boy misses more and more repayments Michael threatens to break his legs in retaliation. Charlie convinces Michael to reduce the extortionate interest and promises to straighten Johnny Boy out. However, an increasingly erratic and irrational Johnny Boy doesn't cooperate and ends up insulting a very angry Michael and then pulls a gun on him. Charlie panics and flees the scene with Johnny Boy in a car, collecting Teresa on the way. Just when they start to relax, a car roars up beside them with Michael and a gunman who fires into Charlie's car – hitting Johnny Boy gruesomely in the neck and Charlie in the arm. Charlie crashes the car, Johnny Boy stumbles clear and collapses, Teresa lies prone in the front seat. Depressingly and tragically, the film ends with the emergency services coming to their aid as Charlie kneels by the wreck in shock.

BIBLIOGRAPHY

Academy of Motion Picture Arts and Sciences. (2006). Retrieved from http://www.oscars.org

Agosti, M. (2002). Information retrieval on the web. In *Proceedings of the European Summer School in Information Retrieval 2000, Lectures on Information Retrieval* (pp. 242-285). Springer.

Bordwell, D., Staiger, J., & Thompson, K. (1985). *The Classical Hollywood Cinema.* Columbia University Press.

Bradley, M. M., Cuthbert, B. N., & Lang, P. J. (1993). Pictures as pre-pulse: attention and emotion in startle modification. *Psychophysiology, 30,* 541-545.

Brady, S., Diamond, D., Lau, K.-T. (2005).Inherently conducting polymer modified polyurethane smart foam for pressure sensing. *Sensors and Actuators a : Physical, 199*(2), 398

Cacioppo, J. , Bernston, G., Larson, J., Poehlmann, K., & Ito, T. (2000). The Psychophysiology of Emotion. In M. Lewis & J. Haviland-Jones (Eds.), *Handbook of Emotions* (2nd ed., pp. 173-191). Guilford Press, New York.

Cognitive Science Laboratory, Princeton University. (2006). *WordNet – A Lexical Database for the English Language.* Retrieved from http://wordnet.princeton.edu

Cytowic, R. E. (1976). Aphasia in Maurice Ravel. *Bulletin of the Los Angeles Neurological Societies, 41*, 109-114.

Fellbaum, C. (Ed.). (1998). *WordNet: An Electronic Lexical Database (Language, Speech, and Communication).* MIT Press.

Frey, W. H. (1985) *Crying: the mystery of tears.* Minneapolis: Winston Press Inc.

Gorbman, C. (1987). *Unheard Melodies.* Indiana Press.

GroupLens Research. (2006). *MovieLens.* Retrieved from http://movielens.umn.edu/login

Lavy, M. M. (2001). *Emotion and the Experience of Listening to Music: A Framework for Empirical Research.* Ph.D. Dissertation, Cambridge, U.K.

Lehane, B. (2006). *Automatic Indexing of Video Content via the Detection of Semantic Events.* Ph.D. Dissertation, Dublin City University, Ireland.

Lehane, B., O'Connor, N., & Lee H. (2006). Searching Movies Based on User Defined Semantic Events. In *Proceedings of the International Conference of Signal Processing and Multimedia Applications (SIGMAP).* Portugal.

Lehane, B. & O'Connor, N. (2006). Movie Indexing via Event Detection. *In Proceedings of the Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS).* Korea.

Lehane B., O'Connor, N., & Murphy, N. (2005). Dialogue Sequence Detection in Movies. In *Proceedings of the International Conference on Image and Video Retrieval (CIVR).* Singapore.

Lehane, B., O'Connor, N., & Murphy, N. (2004a). Dialogue Scene Detection in Movies using Low and Mid-Level Visual Features. In *Proceedings of the International Workshop on Image, Video and Audio Retrieval and Mining.* Quebec, Canada.

Lehane, B., O'Connor, N., & Murphy, N. (2004b). Action Sequence Detection in Motion Pictures. In *Proceedings of the European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology (EWIMT).* London, U.K.

O'Gorman, D. (2005). *Personal Communication: May 2005.* Centre for Sports Science and Health, Dublin City University.

Panksepp, J. & Bernatzky, G. (2002). Emotional Sounds and the Brain: the neuro-affective foundations of musical appreciation. *Behavioural Processes, 60*(2), 133-155.

Pickering, M.J., Wong, L. & Ruger, S. (2003). ANSES: Summarisation of News Video. In *Proceedings of the International Conference on Image and Video Retrieval (CIVR),* University of Illinois, USA.

Rickard, N. S. (2004). Intense emotional responses to music: a test of the physiological arousal hypothesis. *Psychology of Music, 32*(4), 371-388.

Rimland, B. & M. Edelson, S. (1995).A pilot study of auditory integration training in autism. *Journal of Autism and Developmental Disorders, 25*, 61-70.

Rothwell, S., Lehane, B., Chan, C.-H., Smeaton, A. F., O'Connor, N. E., Jones, G. J.F. & Diamond, D. (2006). The CDVPlex Biometric Cinema: Sensing Physiological Responses to Emotional Stimuli in Film. In *Adjunct Proceedings of the 4th International Conference on Pervasive Computing.* Dublin, Ireland.

Russell, J. & Mehrabian, A. (1977). Evidence for a Three-Factor Theory of Emotions. In *Journal of Research in Personality, 11*, 273-294.

Sadlier, D., O'Connor, N., Murphy, N., & Marlow, S. (2004). A framework for event detection in field-sports video broadcasts based on SVM generated audio-visual feature model. Case-study: soccer video. In *Proceedings of the International Workshop on Systems, Signals and Image Processing (IWSSIP).*

Salway, A. & Graham, M. (2003). Extracting Information about Emotions in Films. In *Proceedings of the 11th ACM Conference on Multimedia* (pp. 299-302). California, U.S.A.

Selfridge-Field, E. (1998). Conceptual and representational issues in melodic composition. In W.B. Hewlett & E. Selfridge-Field (Eds.), *Melodic Similarity – Concepts, Procedures and Applications.* MIT Press, Cambridge, Massachusetts, USA.

Soifer, R.(1997). *Music in Video Production.* Focal Press.

Sonnenschein, D. (2001). *Sound Design.* Michael Wiese Productions.

The Centre for Digital Video Processing (2006). Retrieved from http://www.cdvp.dcu.ie

The Four Elements of Music. (2005). Retrieved from http://www.pianomusic4you.com/elementsofmusic.html

The Internet Movie Database. (2006). Retrieved from http://www.imdb.com

Thomas, T. (1997). *Music for the Movies* (2nd Ed). Silman-James Press.

Varian, H. & Lyman, P. (2003). *How much information 2003.* Retrieved from www.sims.berkeley.edu/research/projects/how-much-info-2003

Walker, R. (2004). Cultural Memes, Innate Proclivities and Musical Behaviour: A Case Study of the Western Traditions. In *Psychology of Music, 32*(2), 153-190.

WEBTEH Ltd. (2006). *BS.Player.* Retrieved from http://www.bsplayer.org