

Methods for Affective Content Analysis and Recognition in Film

S A ROBERTS

PhD 2022

Methods for Affective Content Analysis and Recognition in Film

SHAUN ANDREW ROBERTS

A thesis submitted in partial fulfilment of the
requirements of Manchester Metropolitan
University for the Degree of Doctor of
Philosophy

Department of Computing and Mathematics
Manchester Metropolitan University

2022

Statement of Originality


I, Shaun Roberts, confirm that the research included within this thesis is my work or that where it has been carried out in collaboration with or supported by others, that this is duly acknowledged below, and my contribution is indicated.

I attest that I have exercised reasonable care to ensure that the work is original and, to the best of my knowledge, does not break any UK law, infringe on any third party's copyright or other intellectual property rights, or contain any confidential material.

I accept that the university has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author, and no quotation from it or information derived from it may be published without the author's prior written consent.

Signature: 

Date: Friday, 13 January 2023

Abstract

The research presented in this thesis resulted from the growing attention on the effects of emotion on users, raising questions about their potential application to computational systems.

This research investigates the best methods for determining affective scoring for video content, specifically films. This resulted in the affective video system (AVS) framework, AVS dataset and AVS systems being developed, leading to several contributions to knowledge about the best affective methods and systems.

This work presents the necessary theory to understand the subject area. It builds as the thesis matures, laying a pathway in the form of a methodology framework for viewing affective problems and systems, moving into a subsequent study reviewing the well-recognised affective methods such as the International Affective Picture System (IAPS) and how its well-defined processes and procedures could be adapted for a more modern approach using video content. The research then studies the most critical perceivable features from video clips for users, which were analysed using the repertory grid approach.

This led to the above contributions being combined to create the AVS system and database, which is a unique database comprising the affective scores for various film clips. This research concluded with the presentation of the best regression methods resulting from this research and its datasets and a summary of this performance, and discussions of the results in terms of other research in this area.

Table of Contents

ABSTRACT.....	5
TABLE OF CONTENTS.....	6
LIST OF FIGURES.....	14
LIST OF TABLES.....	17
ACKNOWLEDGEMENTS.....	18
CHAPTER 1 - INTRODUCTION.....	19
1.1 PROBLEM STATEMENT.....	20
1.2 RESEARCH AIMS.....	21
1.3 THESIS STRUCTURE.....	22
1.3.1 CHAPTER 1 - INTRODUCTION.....	22
1.3.2 CHAPTER 2 - LITERATURE REVIEW.....	22
1.3.3 CHAPTER 3 - STRATEGY AND RESEARCH METHODS.....	22
1.3.4 CHAPTER 4 - AN INTERNATIONAL AFFECTIVE PICTURE SYSTEM (IAPS) APPROACH FOR AFFECTIVE VIDEO SYSTEM (AVS) A RATING METHODOLOGY FOR AFFECTIVE VIDEO.....	23
1.3.5 CHAPTER 5 - PERCEPTIONS OF VIDEO FEATURES USING A REPERTORY GRID.....	23
1.3.6 CHAPTER 6 - AVS DATASET CREATION, AVS DATA ANALYSIS.....	23
1.3.7 CHAPTER 7 - PREDICTING AFFECTIVE RESPONSES FOR VIDEO CONTENT USING MACHINE LEARNING.....	24
1.3.8 CHAPTER 8 - CONCLUSION.....	24
1.4 BACKGROUND TO AFFECTIVE COMPUTING.....	25
1.5 CHALLENGES TO AFFECTIVE COMPUTING.....	26
1.5.1 CHALLENGES IN SENSING AND RECOGNISING EMOTIONS.....	26
1.5.2 CHALLENGES AROUND AFFECT MODELLING.....	28
1.5.3 CHALLENGES AROUND EMOTIONAL EXPRESSION.....	28
1.5.4 ETHICAL CHALLENGES.....	29

CHAPTER 2 - LITERATURE REVIEW	30
2.1 MODELLING AFFECT	30
2.1.1 EMOTIONS AS EXPRESSIONS	30
2.1.2 EMOTIONS AS EMBODIMENTS	31
2.1.3 COGNITIVE APPROACHES TO EMOTIONS	31
2.1.4 HOW EMOTIONS INFORM JUDGMENT AND REGULATE THOUGHT	31
2.1.5 EMOTIONS AS SOCIAL CONSTRUCTS	32
2.1.6 IDENTIFYING THE PHYSICAL MANIFESTATIONS OF EMOTION	33
2.1.6.1 ELECTRODERMAL ACTIVITY (EDA)	33
2.1.6.2 EDA SETUP	34
2.1.6.3 BODY LANGUAGE	35
2.1.6.4 FACIAL RECOGNITION	35
2.1.7 PERSONALISED AFFECTIVE RECOMMENDATIONS	36
2.1.8 CRITERIA FOR EMOTIONAL RECOGNITION	36
2.1.9 EMOTION CLASSIFICATION SYSTEMS	38
2.1.10 CIRCUMPLEX MODEL	40
2.1.11 GENEVA EMOTION WHEEL (GEW)	43
2.1.12 PLUTCHIK'S MODEL	43
2.1.13 SELF-ASSESSMENT MANIKIN (SAM)	45
2.1.14 IAPS	46
2.1.15 LIRIS-ACCEDE	49
2.1.16 FILM STIM	50
2.1.16.1 VIDEO CONTENT DESIGN	51
2.1.16.2 EXPERIMENTAL DESIGN	52
2.1.16.3 EXPERIMENTAL MEASURES	52
2.1.17 THE EMOTIONAL MOVIE DATABASE (EMDB)	53
2.2 RECOMMENDATION SYSTEMS	54
2.2.1 COLLABORATIVE FILTERING (CF)	55
2.2.2 BASELINE PREDICTORS	56

2.2.3 USER–USER COLLABORATIVE FILTERING (CF).....	56
2.2.4 CONTENT-BASED FILTERING.....	58
2.2.5 KNOWLEDGE-BASED RECOMMENDER SYSTEM	58
2.2.5.1 CASE-BASED RECOMMENDATIONS.....	59
2.2.5.2 CONSTRAINT-BASED RECOMMENDATIONS.....	59
2.2.6 HYBRID RECOMMENDER SYSTEMS.....	59
2.2.7 SUMMARY OF RECOMMENDER SYSTEMS.....	59
2.2.8 MACHINE LEARNING	60
2.2.9 REGRESSION	61
2.3 RELATED WORK.....	62
2.3.1 EMOTIONAL CONTEXT AWARENESS	62
2.3.2 AFFECTIVE METHODS FOR TV SERVICES.....	64
2.3.3 AN EVALUATION OF THE INTERNATIONAL AFFECTIVE PICTURE SYSTEM.....	65
2.3.4 IFELT: ACCESSING MOVIES THROUGH OUR EMOTIONS.....	67
2.3.5 MINING AFFECTIVE CONTEXT IN SHORT FILMS FOR EMOTION-AWARE RECOMMENDATION	68
2.4 SUMMARY.....	68
CHAPTER 3 - AFFECTIVE VIDEO SYSTEM FRAMEWORK (AVSF)	70
3.1 AVSF OVERVIEW	70
3.2 AVSF FRAMEWORK OVERVIEW	71
3.3 AFFECTIVE INPUTS.....	72
3.4 AFFECTIVE PROCESSING.....	73
3.5 AFFECTIVE AWARE OUTPUTS.....	73
3.6 CONCLUSION	74
CHAPTER 4 - AN INTERNATIONAL AFFECTIVE PICTURE SYSTEM (IAPS) APPROACH FOR AFFECTIVE VIDEO SYSTEM (AVS) A RATING METHODOLOGY FOR AFFECTIVE VIDEO.....	75
4.1 INTRODUCTION	75

4.2 AIMS.....	78
4.3 ANALYSIS METHODS	78
4.3.1 PARTICIPANTS	78
4.3.2 EXPERIMENTAL OVERVIEW	79
4.3.3 EXPERIMENTAL PROCEDURE	81
4.3.3.1 VIDEO CONTENT.....	83
4.3.3.2 RATING PROCEDURE	83
4.3.3.3 CONTENT CLASSIFICATION	84
4.4 RESULTS	85
4.4.1 AVS DATASET	87
4.4.2 PAIRED T-TEST	87
4.4.3 PEARSON CORRELATION RESULTS.....	87
4.4.4 AFFECTIVE SPACE COMPARISON	89
4.4.5 IAPS RESULTS.....	90
4.4.6 AVS RESULTS	91
4.5 DISCUSSION.....	93
4.6 CONCLUSION	95
4.7 SUMMARY.....	96
4.8 PROGRESSION	96
CHAPTER 5 - PERCEPTIONS OF VIDEO FEATURES USING A REPERTORY GRID	97
5.1 INTRODUCTION	97
5.2 AIMS.....	98
5.3 REPERTORY GRID	98
5.4 ANALYSIS METHODS	100
5.4.1.1 REPERTORY GRID METHOD OVERVIEW	100
5.4.1.2 EXAMPLE REPERTORY GRID AS DRAWN ON A WHITEBOARD	101
5.4.1.3 REPERTORY GRID EXPERIMENTAL PROCEDURE	103

5.4.2 PARTICIPANTS	104
5.4.3 EXPERIMENTAL METHODOLOGY	105
5.4.4 WEB INTERFACE	106
5.5 RESULTS	107
5.5.1 PRINCIPAL COMPONENT GRID	108
5.5.2 FILM CLIP RESULTS	109
5.5.3 CONSTRUCT ANALYSIS.....	110
5.5.4 CONSTRUCT ANALYSIS.....	112
5.5.5 WORD ANALYSIS.....	114
5.6 DISCUSSION.....	114
5.7 CONCLUSION	115
5.8 SUMMARY.....	116
5.9 PROGRESSION	117
CHAPTER 6 - AVS DATASET CREATION, AVS DATA ANALYSIS	118
6.1 INTRODUCTION	118
6.2 AIMS.....	119
6.3 METHODS.....	120
6.3.1 MATERIALS	120
6.3.2 PARTICIPANTS	123
6.3.3 RATING MECHANISM (THE SAM SCALE).....	126
6.3.4 EXPERIMENT DESIGN.....	127
6.3.4.1 TESTING OVERVIEW	129
6.3.4.2 EXPERIMENT OVERVIEW	130
6.4 RESULTS	131
6.4.1 AVS DESCRIPTIVE STATISTICS	131
6.4.2 AVS FILM CLIPS RESULTS	132
6.4.3 DISTRIBUTION OF RATINGS	132

6.4.4 COMPARISON TO OTHER DATASETS	135
6.5 DISCUSSION.....	135
6.5.1 CONTENT LENGTH	135
6.5.2 CONTENT	136
6.5.3 NUMBER OF RATINGS	137
6.5.4 LIMITATIONS	137
6.6 CONCLUSION	137
6.7 SUMMARY.....	139
6.8 PROGRESSION	139
 CHAPTER 7 - PREDICTING AFFECTIVE RESPONSES FOR VIDEO CONTENT USING MACHINE	
LEARNING.....	140
7.1 INTRODUCTION	140
7.2 AIMS.....	141
7.3 METHODS.....	142
7.4 FEATURE EXTRACTION	142
7.4.1 QCTOOLS (QUALITY CONTROL TOOLS FOR VIDEO PRESERVATION).....	142
7.4.2 RGB (RED, GREEN, AND BLUE).....	143
7.4.3 AUDIO TOOLBOX MATLAB.....	143
7.4.4 MUSIC INFORMATION RETRIEVAL (MIR) TOOLBOX.....	143
7.4.5 MATLAB FEATURE EXTRACTION	143
7.5 EXTRACTED FEATURES	143
7.5.1.1 AUDIO FEATURES	144
7.5.1.2 VIDEO FEATURES.....	144
7.6 MACHINE LEARNING TECHNIQUES.....	145
7.6.1 REGRESSION MODELS	145
7.6.2 CROSS-VALIDATION.....	146
7.6.3 MODEL PERFORMANCE METRICS.....	146

7.6.3.1 MEAN SQUARED ERROR (MSE)	146
7.6.3.2 ROOT MEAN SQUARED ERROR (RMSE)	147
7.6.3.3 R-SQUARED (R^2)	148
7.6.3.4 MEAN ABSOLUTE ERROR (MAE).....	149
7.6.3.5 PREDICTION SPEED (~OBS/SEC)	150
7.7 RESULTS	150
7.7.1.1 MODEL PERFORMANCE METRICS	151
7.7.1.2 THE TOP 10 BEST PERFORMING MODELS (RMSE) CROSS-5 – AROUSAL	152
7.7.1.3 THE TOP 10 BEST PERFORMING MODELS (RMSE) CROSS-10 – AROUSAL	153
7.7.1.4 THE TOP 10 BEST PERFORMING MODELS (RMSE) CROSS-5 – VALENCE.....	154
7.7.1.5 THE TOP 10 BEST PERFORMING MODELS (RMSE) CROSS 10 – VALENCE	155
7.8 DISCUSSION.....	157
7.8.1 DISCUSSION OF 5-FOLD CROSS-VALIDATION AROUSAL	158
7.8.2 DISCUSSION OF 5-FOLD CROSS-VALIDATION VALENCE	162
7.9 CONCLUSION	165
7.10 SUMMARY.....	166
CHAPTER 8 - CONCLUSION.....	168
8.1 ORIGINAL RESEARCH QUESTIONS	169
8.1.1 WHAT CHANGES WOULD NEED TO BE MADE TO THE INTERNATIONAL PICTURE SYSTEM (IAPS) TO PROVIDE AFFECTIVE SCORING FOR VIDEO	169
8.1.2 WHAT VIDEO FEATURES ARE MOST IMPORTANT TO THE USERS IN THEIR PERCEPTION OF VIDEO, AND FROM THESE FEATURES, WHICH CAN BE USED IN MACHINE LEARNING REGRESSION METHODS?	169
8.1.3 HOW CAN A DATASET BE CREATED FOR VIDEO WHICH CAN BE UTILISED IN THE SAME WAY AS THE IAPS?	170
8.1.4 WHAT FEATURES CAN BE EXTRACTED FROM VIDEO CONTENT THAT CAN BE USED TO PREDICT AFFECT?	170
8.1.5 WHICH LEARNING REGRESSION MODELS WILL BEST PREDICT AFFECTIVE SCORES FOR VIDEO CONTENT?.....	170
8.2 CONTRIBUTIONS.....	171
8.3 FUTURE RESEARCH: AFFECTIVE FEATURES.....	175

REFERENCES	177
------------------	-----

List of Figures

FIGURE 2:1 - EDA SETUP (PFLANZER & McMULLEN, 2000).....	35
FIGURE 2:2 - FOUR-WAY BINARY CLASSIFICATION (MOWER, ET AL., 2011)	38
FIGURE 2:3 - EXAMPLE OF A DIMENSIONAL AFFECTIVE MODEL (ZHANG, ET AL., 2010).....	39
FIGURE 2:4 - CIRCUMPLEX MODEL (RUSSELL, 1980)	40
FIGURE 2:5 - ALTERNATIVE DIMENSIONAL STRUCTURES OF THE SEMANTIC SPACE FOR EMOTIONS (SCHERER, 2005)	41
FIGURE 2:6 (A) CIRCUMPLEX MODEL OF AFFECT, (B) MODEL SIMPLIFICATION THROUGH DISCRIMINATION OF FOUR CATEGORIES (NIESE ET AL., 2011)	42
FIGURE 2:7 - FACIAL EXPRESSION MAPPED TO EMOTIONAL OUTCOME CHART (NIESE ET AL., 2011).....	42
FIGURE 2:8 - VERSION 2.0 OF THE GEW WITH 40 EMOTION TERMS ARRANGED IN 20 EMOTION FAMILIES (SACHARIN, 2012).....	43
FIGURE 2:9 PLUTCHIK'S MODEL (PLUTCHIK, 1997)	44
FIGURE 2:10 SAM SCALE DIMENSIONS OF VALENCE (TOP), AROUSAL (MIDDLE), AND DOMINANCE (BOTTOM) (MEHRABIAN & RUSSELL, 1974).	46
FIGURE 2:11 NORMALISED DISTRIBUTION OF FILMS BY GENRE INCLUDED IN LIRIS-ACCEDE (BAVEYE, ET AL., 2015).....	50
FIGURE 2:14 - EXAMPLE OF THE DEGREE OF EMOTIONAL SCALING (YOO, ET AL., 2011)	63
FIGURE 2:15 - AFFECTIVE BENCHMARKING SOLUTION APPLIED TO MOVIES (FLEUREAU, ET AL., 2013).	65
FIGURE 2:16 - SCATTER DIAGRAM OF PICTURES BY AROUSAL AND VALENCE DIMENSIONS (TOK, 2010).	67
FIGURE 3:1 AVS FRAMEWORK	74
FIGURE 4:1 – AVS TESTING PROCESS	80
FIGURE 4:2 - AVS OVERVIEW	81
FIGURE 4:3 - THE SELF-ASSESSMENT MANIKIN (BRADLEY & LANG, 1994).	83
FIGURE 4:4 - FOUR-WAY BINARY CLASSIFICATION (MOWER ET AL., 2011)	84
FIGURE 4:5 - EXAMPLE OF HOW SAM SCORES MAP TO THE CIRCUMPLEX MODEL	85
FIGURE 4:6 - GRAPHIC REPRESENTATION OF DESCRIPTIVE STATISTICS FOR THE IAPS VS AVS	86
FIGURE 4:7 – PEARSON CORRELATION GRAPH FOR AROUSAL	88
FIGURE 4:8 - PEARSON CORRELATION GRAPH FOR VALENCE	89
FIGURE 4:9 – OVERVIEW OF AFFECTIVE SPACE RESULTS BETWEEN THE IAPS AND AVS.....	90

FIGURE 4:10 - IAPS RESULTS	91
FIGURE 4:11 - AVS RESULTS	92
FIGURE 4:12 - OVERLAY VALUES FOR IAPS AND AVS. THE AFFECTIVE SPACE HAS BEEN GAUGED ON AROUSAL AND VALENCE DIMENSIONS.	94
FIGURE 5:1 - EXAMPLE REPERTORY GRID FROM EXPERIMENT	101
FIGURE 5:2 - CARD'S EXAMPLE (CUT UP INDIVIDUALLY)	102
FIGURE 5:3 – THE SYSTEM USED TO PRESENT VIDEO CONTENT TO THE USER.	106
FIGURE 5:4 - PRINCIPAL COMPONENT GRID MAP	108
FIGURE 5:5 – PERCENTAGE OF SIMILARITY BETWEEN CONTENT	109
FIGURE 5:6 – FOCUS CLUSTER DENDROGRAM FOR PERSONAL CONSTRUCTS	111
FIGURE 5:7 - MOST RECURRING CONSTRUCTS	114
FIGURE 6:1 - AVS DATASET GENRE COMPOSITION	122
FIGURE 6:2 - AVS DATA COLLECTION LOCATIONS OVERVIEW	124
FIGURE 6:3 - AVS DATA COLLECTION LOCATIONS HEATMAP	124
FIGURE 6:4 - OVERVIEW OF PARTICIPATION BY COUNTRY	124
FIGURE 6:5 - GENDER OVERVIEW	125
FIGURE 6:6 - AGE OVERVIEW	125
FIGURE 6:7 – EXPERIMENTAL PROCEDURE OVERVIEW	128
FIGURE 6:8 - AVS TESTING OVERVIEW	129
FIGURE 6:9 - EXAMPLE OF HOW USERS WERE PRESENTED THE CLIPS DURING THE AVS STUDY	130
FIGURE 6:10 - AN AFFECTIVE OVERVIEW OF THE VIDEO CONTENT	131
FIGURE 6:11 - ALL CLIPS AVERAGE AROUSAL AND VALENCE ON SAM SCALE	133
FIGURE 6:12 - ALL MOVIES' AROUSAL AND VALENCE POINT ON A FOCUSED GRAPH	134
FIGURE 7:6 - 5-FOLD CROSS-VALIDATION - AROUSAL TRUE Vs PREDICTED (CUBIC SVM)	159
FIGURE 7:7 - 5-FOLD CROSS-VALIDATION - AROUSAL PREDICTED VS ACTUAL (CUBIC SVM)	160
FIGURE 7:8 - 5-FOLD CROSS-VALIDATION - AROUSAL RESIDUALS (CUBIC SVM)	161
FIGURE 7:9 – 5-FOLD CROSS-VALIDATION—VALENCE TRUE Vs PREDICTED (LINEAR SVM)	162

FIGURE 7:10 - 5-FOLD CROSS-VALIDATION - VALENCE PREDICTED VS ACTUAL (LINEAR SVM)	163
FIGURE 7:11 – 5 -FOLD CROSS-VALIDATION - VALENCE RESIDUALS (LINEAR SVM)	164
FIGURE 7:12 - AVS COMPARED WITH LIRIS PROTOCOL A RESULTS	165

List of Tables

TABLE 3:1 AVSF INPUT, PROCESS AND OUTPUT APPROACH	72
TABLE 4:1 – STATISTICAL OVERVIEW TO PARTICIPANTS OF THE STUDY	79
TABLE 4:2 – DESCRIPTIVE STATISTICS FOR THE IAPS VS AVS.....	85
TABLE 5:1 - FILM CLIPS INCLUDED IN THE REPERTORY GRID STUDY WITH IMDB GENRES (IMDb, 2021)	107
TABLE 5:2 - 100% MATCH FOR CONSTRUCTS.....	112
TABLE 5:3 - 96.9% MATCH FOR CONSTRUCTS.....	113
TABLE 5:4 - 96.9% MATCH FOR CONSTRUCTS.....	113
TABLE 6:1 - DESCRIPTIVE STATISTICS FOR THE AVS DATA	132
TABLE 6:2 - SIMILAR STUDIES OVERVIEW	135
TABLE 7:1 - TOP 10 BEST PERFORMING MODELS (RMSE) FOR CROSS 5 – AROUSAL	152
TABLE 7:2 - TOP 10 BEST PERFORMING MODELS (RMSE) FOR CROSS-10 – AROUSAL	153
TABLE 7:3 - TOP 10 BEST PERFORMING MODELS (RMSE) FOR CROSS 5 – VALENCE	154
TABLE 7:4 - TOP 10 BEST PERFORMING MODELS (RMSE) FOR CROSS 10 – VALENCE	155
TABLE 7:5 - DIRECT COMPARISON OF THE TOP 3 REGRESSION METHODS.....	157

Acknowledgements

First and foremost, I would like to acknowledge and thank my principal supervisor, Stuart Cunningham, for his time and helpful guidance throughout the course of the research, and above all for his patience and understanding. I would also like to acknowledge and again thank my first supervisor, John Darby, for his time and helpful supervision. In addition, I would like to thank Jon Weinel, Jason Matthews, and Paul Comerford for their support and advice.

Lastly, I would like to thank my family and friends for their continuing support during the research and writing of this thesis.

Chapter 1 - Introduction

This research resulted from the increasing interest in the effects of emotion on users, which has created the need for reliable techniques for emotional identification and classification. The term 'affective computing' was introduced by Rosalind Picard to describe computing methods that relate to, arise from, or deliberately influence emotion or other affective phenomena (Picard & Healey, 1997). The term also encompasses a computational model of human emotion in implementing autonomous agents capable of affective processing (Scherer, et al., 2010).

This research will review current methods and approaches utilised in affective computing, and video content analysis to provide an alternative to the primarily based collaborative filtering/genre approach currently adopted. The theory behind this research is that users could (in the future) browse content recommendations based on the emotional experience they seek. Users now have hundreds of thousands of videos available across a vast network. This mass of multimedia content has stretched the current recommendation system methods, which means that more and more resources are dedicated to improving the system's recommendation processes.

This research comes at a time after the role of emotions across humanity has been recognised as a driving factor for decision-making, contradicting earlier theories that suggested a logical approach to decision-making, which has now been largely discredited. Instead, prevailing theories suggest that users can plausibly explain their thinking and

behaviour. Still, in reality, 95% of thought occurs in our unconscious minds, and people use consciousness to rationalise behaviour (Zaltman, 2003).

1.1 Problem statement

Recommendation systems for film are limited in efficiency, especially when working with large content databases. Integrating affective information into the recommendation process may be one way to enhance the process. However, there currently needs to be more real-world affective studies in film that could help improve recommendation systems in future. This would be achieved by including affective data in the decision-making process. Picard outlines the potential implications for these systems and how they may become closer to human intelligence with the addition of affective data (Picard, et al., 2001)

The problem is also compounded by the sheer amount of multimedia data, which makes showing the correct content to users much harder (Yazdani, et al., 2013). There have always been issues around understanding human emotion and its impact, leading to research such as (Russell, 1980) to construct theories to help explain human emotion.

As affective research grew, the application of the research grew also. Picard presented research linking emotion with machines (Picard, et al., 2001). This opened the door to exploring how affective data could be used to improve (in this context) machine learning processes that utilise affective data.

The research aims to bridge the gap between the user's affective data for video content in the form of films. This research can be used in subsequent studies to define the best way to move forward and develop this technology into real-world applications, such as providing

users with additional categorisation options when selecting video content based on its affective score, thus moving beyond purely genre-based methods.

1.2 Research aims

The foremost concern of this thesis is to investigate the best methods for determining affective scoring for video content, specifically films. The intention drives this is that such knowledge could be used in future to improve decision-making processes in recommendation and categorisation systems, by way of a real-world application. This led to formulating research questions about the systems and techniques needed to underpin it, which are explored in this research.

This thesis addresses the following research questions:

1. What changes would need to be made to the International Affective Picture System (IAPS) to provide an equivalent affective scoring approach for video?
2. What video features are most important to the users in their perception of video, and from these features, which can be used in machine learning regression methods?
3. How can a dataset be created for video which can be utilised in the same way as the IAPS?
4. What features can be extracted from video content that can be used to predict affect?
5. Which machine learning regression models will best predict affective scores for video content?

1.3 Thesis structure

1.3.1 Chapter 1 - Introduction

Chapter 1 introduces this thesis, outlining the research objectives and the rationale behind the research, highlighting research relevancy, and what research the thesis will present.

1.3.2 Chapter 2 - Literature Review

Chapter 2 evaluates all the literature relevant to an 'affective video system (AVS)'. The literature review covers concepts from numerous fields, such as affective computing, machine learning, and psychology. These are analysed for their significance to the current research. Subsections start with an overview of these topics and then explore their relevance to an AVS in more depth.

Applications and approaches to affective computing are reviewed, and the elements that underpin their design are highlighted. Subsequently, this literature review section synthesises the findings from the literature into a unique approach.

1.3.3 Chapter 3 -

Chapter 3 synthesises the literature research in Chapter 2 and builds on it to propose a research approach for an AVS that encompasses the theory behind affective methods, how to interpret affect, and how the collected affective data can be used to improve current methods of content categorisation and content recommendations.

This framework is fundamental as there are several relevant technologies and research topics that include affective computing, IAPS, and machine learning, which must be

combined effectively to achieve an AVS. The affective video system framework (AVSF) helps break down these systems and allows specific research questions to be investigated.

1.3.4 Chapter 4 - An International Affective Picture System (IAPS) Approach for Affective Video System (AVS) A Rating Methodology for Affective Video

Chapter 4 investigates whether the affective scoring of content based on its IAPS counterpart is comparable. This would mean that a well-recognised pre-existing method could form the basis of the new system, ensuring the affective data collected is statistically significant to the pre-existing IAPS methodology.

1.3.5 Chapter 5 - Perceptions of Video Features Using a Repertory Grid

Chapter 5 identifies perceptual features' importance to users from film trailers using the repertory grid methodology. As repertory grids are used to understand relationships between constructs and a series of elements of interest, this method produces implicit theory about stimuli the users are exposed to. The study also provides an overview of the key visual and audio features perceived in the content of various film genres. This will lead to the formation of a generalised group view of the film clips presented.

1.3.6 Chapter 6 - AVS Dataset Creation, AVS Data Analysis

Chapter 6 describes an experiment examining the role of affective responses to film trailers presented via the AVS system, which allows for affective data to be gathered directly from the users after viewing video content. The AVS dataset, which is comprised of 100 clips ranging from 60 to 226 seconds and 6202 affective scores based on the SAM scale for arousal and valence, is also presented in this chapter. Chapter 7 includes the original and

unique affective dataset presented in this study, which gives a much clearer insight into film trailers' affective scoring.

1.3.7 Chapter 7 - Predicting Affective Responses for Video Content Using Machine Learning

Chapter 7 outlines a method for affective video content analysis, including how to extract audio and video features from the video content, which were used in conjunction with the AVS dataset covered in Chapter 7. Regression learning methods were then presented to see which of the regression methods yielded the best results.

1.3.8 Chapter 8 - Conclusion

Chapter 8 presents the overarching conclusions of this research, summarising its contributions to date. Finally, section 8.3, future research, suggests potential future work for moving this research forward, highlighting several key areas, and further studies that could be conducted to develop and expand upon the work presented in this thesis.

1.4 Background to Affective Computing

Affective computing is the name for computing methods that relate to the influence of emotions or other phenomena upon the user. The term refers directly to the knowledge of a given feeling/emotion. For example, it could be used to process video content in recommendation systems because the decisions need to be directly relevant to the user (Picard & Healey, 1997).

The term 'affective computing' was coined by Rosalind Picard and has been accepted as the label for computing methods that relate to, arise from, or deliberately influence emotion or other affective phenomena (Picard & Healey, 1997). Furthermore, the term affective computing is widely used to refer to the computational model of human emotion within the implementation of autonomous agents capable of affective processing (Scherer, et al., 2010).

Affective computing relates to the experience of a given feeling or emotion. It could play a crucial part in processing the user context because the decisions need to be directly relevant to the user. By taking affective elements into account, the system may be able to understand user preferences and fine-tune decisions appropriately.

Picard's three categories were:-

- Computers that recognise emotions
- Computers that pretend to have emotions
- Computers that do have emotions.

Of these three, the most important is the recognition of emotions in terms of their contribution to this research. These systems respond to a user's emotions with emotionally supportive interactions, demonstrating components of human awareness. Furthermore, the working systems support the prediction that a computer can begin to understand and react to this information (Picard & Klein, 2002).

1.5 Challenges to affective computing

Picard outlines some of the challenges this technology faces (Picard, 2003), as well as the benefits of adding an emotional component to computers and the inherent benefits that this could provide.

1.5.1 Challenges in sensing and recognising emotions

Challenges that have arisen in affective computing as the technology has developed are outlined in Picard's work (2003). Picard's work covers the complexities of sensing and recognising emotions, as they could be based on factors such as blood chemistry, brain activity, neurotransmitters, and many others, making them non-differentiated. People's expression of emotion is also variable and idiosyncratic, making it highly personal and on a case-by-case basis.

Picard addresses these criticisms by exploring simple narratives such as "if you were asked to jot the emotional state of the next person, you see the challenge grows significantly," highlighting the difference between verbal and emotional expression and perceptual emotion classification. This highlighted that many people do not know how to articulate the

emotional state and the issues while recognising or categorising what they are feeling, and sometimes even assigning a single word to these feelings.

Picard referred to a study that included sensors that monitored an electromyogram, skin conductance, blood volume pulse, and respiration. Also, exploring wearable computers specifically assigned to one person and investigated variations between one person and another, as well as differences between the same person for the same emotion.

The study concluded that the pattern recognition used within this study had an 81% classification accuracy and was not limited to only arousal classifications. The study also highlighted that the same emotions in day-to-day were more significant than variations between emotions on the same day. Picard concluded that physiological differences appear with the eight emotions they used for testing, but it would be incorrect to say that their model could predict emotions with 80% accuracy (Picard, 2003).

“Affect, like weather, is hard to measure, and like weather, it probably cannot be predicted with perfect reliability.” (Picard, 2003, p. 58)

So, in summary, the idea that we can get an idea of some affective states, but not others, was highlighted. Therefore, affective computing will need to grow as a predictive technology, in a similar manner as weather systems, which may not be able to detect with 100% accuracy the weather for a given day but can prepare you by providing you with a forecast for a given day and/or time (Picard, et al., 2001).

1.5.2 Challenges around affect modelling

The criticisms raised around the modelling aspect were that little progress had been made in cognitive modelling, and how it would be possible to achieve something, in theory, more complex.

Picard highlights that the existing models for emotion are, in her view, “stylised stereotypes” of personality types and emotional responsiveness, and that they do not necessarily correspond to actual behaviour in humans. When combined with a lack of understanding of emotional situations and factors, this makes it highly unlikely to be constructed soon.

Picard outlined a research vision that followed a similar pathway as machine research. Even though they do not particularly understand all the intricacies at play, researchers are looking into modelling human vision, and researchers are looking at creating machines. These two fields of research and others can work together and learn from each other to create systems. Picard notes that the systems do not necessarily need to follow nature’s mechanics exactly (Picard, 2003).

1.5.3 Challenges around emotional expression

The criticism of emotional expression is based on computers having no physical body, making it harder for them to express emotions (embodied). Current attempts are currently considered unrealistic and, therefore unconvincing.

Picard claims that computers can have bodies; however, they are different from our own. Referencing the works of Disney and *Star Wars*, she shows that fictional characters can

display their emotions without necessarily conforming to how humans traditionally would convey these emotions (Picard, 2003).

1.5.4 Ethical challenges

Criticisms of the ethics of this type of system were also raised due to the sensitive nature of emotions. Providing information about the most intimate motivational factors could raise ethical issues, potentially leading to unwanted attempts to detect, recognise or manipulate a user's emotional state, leading to a distrust of computers in general.

Picard replied to this criticism by using a real-world example of a boss who may be angry with an employee, where it is ethically wrong to try to alleviate the anger, as it could potentially be viewed as manipulation. She pointed out that humans typically detect, recognise, and respond to emotions or manipulate them in ways that could be considered highly ethical and desirable. For example, playing music to lighten a friend's mood could be considered manipulation, but it is perfectly acceptable (Picard, 2003).

Criticisms raised here are valid and require further consideration, as with any new technology, there is always an unforeseen risk. Whilst the examples provided were generally favourably; there are many examples where this technology could be used for the exact opposite. For example, Picard focused on friends, and it could be argued that this technology could provide a dark advantage in emotional manipulation in the hands of an enemy.

Chapter 2 - Literature Review

This literature review will appraise all the studies relevant to an 'affective video system (AVS).' First, concepts from numerous fields, such as affective computing, machine learning, and psychology, are analysed for their significance to the current research. Subsections start with an overview of these topics and then explore their relevance to an AVS in more depth. Next, applications and approaches to affective computing are reviewed, and the elements that underpin their design are highlighted. Subsequently, this literature review section synthesises the findings from the literature into a unique approach.

2.1 Modelling affect

This section will briefly explore theories derived from traditional emotional theory over several decades while extending theoretical contributions from affective neuroscience. Finally, covering the relevant theories and studies to help inform further research and outlining emotions in relevance to this research for their adoption into an affective system.

2.1.1 Emotions as expressions

Darwin was the first to study emotions. His research recognised a relationship between humans' facial and body expressions and those of other animals. He concluded that emotions resulted from evolutionary processes (Vlachostergiou, et al., 2014). However, Russell suggested that modern evolutionary theory renders Darwin's specific analysis of expressions obsolete (Russell, et al., 2003).

2.1.2 Emotions as embodiments

Russell (1980) proposed a model that combines expressions and physiology and interprets psychological changes as the emotion itself, rather than its expression. The amount of data that can be captured from physiological signals is increasing due to improvements in sensor technology accuracy (see Section 2.1.6).

2.1.3 Cognitive approaches to emotions

Cognitivists believe that for a person to experience a given emotion, an object or an event, it must be considered as something that directly affects them personally, based on that person's experience, goals, and opportunity for action (Vlachostergiou, et al., 2014). This view is based on the Schachter-Singer experiment (Schachter, 1962), which stated that emotion is grounded in two factors: physiological arousal and cognitive label. Therefore, even cognitive theorists argue that (measurable) physiological factors play a role in detecting emotions and are relevant to the development of AVS systems.

2.1.4 How emotions inform judgment and regulate thought

Clore and Huntsinger (2007) completed a study on how emotions inform judgements and regulate thoughts. Their study initially discussed experiments finding influences on moods and emotions and the relationship they share with various kinds of decision. They then moved on to find such influences on cognitive processes. Their study focused on the potential impact of emotions on the decision-making process, as this is one of the challenges faced in context-awareness systems. Finally, they tested the hypothesis that users' emotions impact the decision-making process.

The affect-as-information hypothesis explains both judgement and processing effects by assuming that affect serves as a compelling form of information about value. In the case of judgement, value might be assigned to the object of judgement; in the case of processing, by contrast, value might be assigned to the person's own cognitions and inclinations.

(Clore & Huntsinger, 2007, p. 7).

Specific markers, such as emotions and their correct categorisation, may be paramount to the decisions being made. Their experiments consistently showed that positive affective information promotes, and negative affective information inhibits the cognitive responses that are accessible or dominant in a particular situation. The results highlighted the importance that emotions play in the decision-making process. Such as positive affective results seem to promote a dominant response, and that affect regulates personal versus item-specific processing shown in problem-solving situations.

2.1.5 Emotions as social constructs

The works of Vlachostergiou, et al., (2014) claims that emotions are primarily social constructs, meaning that a social analysis is necessary to understand the relationship between emotions truly. Language is also a vital part of the experience of emotions (Vlachostergiou, et al., 2014). Emotion as a social construct complements the work of (Bazire, 2005), in which all interactions contribute to the social construct of user context. However, just as social constructs for the purposes of informing decision-making, they need to be “embodied” within affective computing.

2.1.6 Identifying the physical manifestations of emotion

When designing an affective system, data representing physiological change are needed for analysis in terms of context awareness to allow for a logical outcome. Emotional states are diverse, including aspects of both a mental state and a physical state. An emotion can be represented as a person's internal dynamics. An emotional experience refers to a time when a person is consciously aware of their emotions and perceives their emotional state (Picard & Healey, 1997).

Picard (1997) outlines some sentic modulation factors. Sentic refers to studies of the waveforms of touch, emotion and music. While Picard's (1997) work is relatively old, it remains the foremost work in this field. Sentic modulation factors that Picard has investigated include facial expression, voice intonation, body language, pupillary dilation, respiration, heart rate, temperature, electrodermal response, and blood pressure (Picard, 1997).

2.1.6.1 Electrodermal activity (EDA)

EDA is a property of the human body that causes continuous variation in the electrical characteristics of the skin. There is an assumption that there is an existing relationship between emotional arousal and sympathetic activity. However, physiological change alone does not identify with any precision which emotions are being experienced. EDA measures the autonomic sympathetic changes, altering sweat and blood flow.

The sympathetic nervous system controls sweat secretion, which EDA measures through changes in electrical conductance. It is well-established that there is a link between arousal

and sweat gland activity (Pflanzer & McMullen, 2000). Skin in humans shows several forms of bioelectric phenomena which are more predominant in the human body's extremities, including fingers, palms of the hands, and soles of the feet.

Galvanic skin resistance (GSR) can be measured when an electrical current is steadily applied between two electrodes and measuring the resistance between them which varies in accordance with the emotional states of humans (Pflanzer & McMullen, 2000).

Galvanic skin potential (GSP) is like GSR. However, the electrodes are connected to a voltage amplifier. As any external current is applied, the voltage is measured between them, and this measurement is referred to as GSP (Pflanzer & McMullen, 2000). When GSR and GSP are combined, they constitute a galvanic skin response.

2.1.6.2 EDA Setup

Figure 2:1 shows how EDA should be set up for the users. EDA sensor is typically placed on the index and middle fingers of the hand.

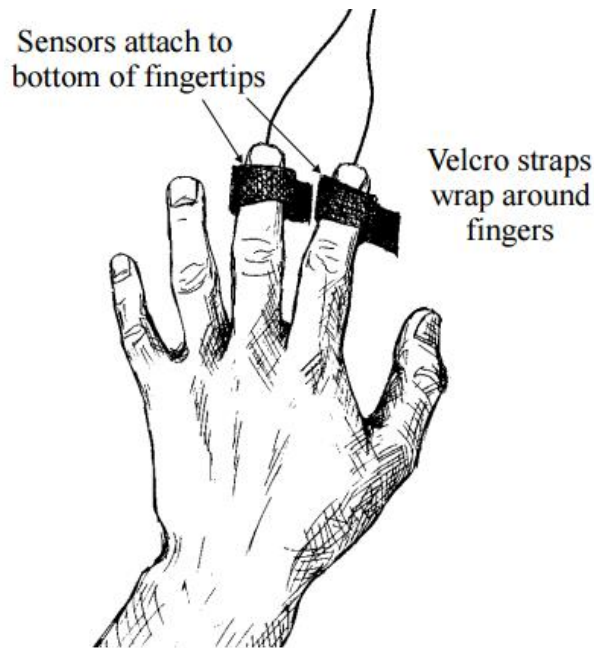


Figure 2:1 - EDA setup (Pflanzer & McMullen, 2000)

2.1.6.3 Body language

Body language is considered an outward reflection of a person's emotional state, where each gesture and/or movement can provide critical insights into the emotion that the individual may be experiencing at a given time. Body language communicates how a person feels without saying anything and, if analysed as part of a set of user attributes, could provide one of the elements for emotional determinations (Pease & Pease, 2016). Body language could be relevant to interpreting user context via affective methods for a recommender system.

2.1.6.4 Facial recognition

Facial recognition is a daily task for humans, but a computational system that can do this as easily as humans have not yet been developed. However, computers can logically interpret

expressions (Jain & Li, 2011). Facial recognition focuses on core facial expressions and numerous micro-expressions. There are two main streams of research into affective recognition from which the face can be identified: facial affect recognition and facial muscle action recognition (Castellano, et al., 2010).

2.1.7 Personalised affective recommendations

One of the significant insights from the research to date has been the subjective nature of an individual's user's personal recommendations. Therefore, once a system is developed to map video content to an affective score correctly, there will be a resulting stage that needs to consider personal preference. This concept is touched upon by (Zhang, et al., 2010) and is referred to as "Personalised Affective Analysis" as an additional step to the recommendation process.

2.1.8 Criteria for emotional recognition

A set of design criteria for emotional recognition within computing, as discussed by (Picard & Healey, 1997), is as follows:

- Input: receives various input signals such as facial expressions, voice, hand gestures, body language, respiration, blood pressure, EDA, temperature, electrocardiogram, pulse, and other physical factors that may contribute to an emotional determination.
- Pattern recognition: full-featured extractions and classifications of the signals.
- Reasoning: predicts potential underlying emotions based on knowledge about how the given option (e.g., happy) was generated via expressions.

- Self-learning: this element is referred to as 'getting to know' the user, learning the individual and their individuality in order to personalise the process, improving accuracy.
- Output: emotion or description to be presented.

Picard's emotional recognition criteria provide a basis for building a system. This framework provides a multitude of techniques to identify emotions, calling on a whole host of different systems to begin deciphering the inner workings behind human emotions. Reasoning in this type of system would be based on logic; the best course of action would be a process of elimination based on using inputs to eliminate several potential emotions.

The self-learning aspect of the system is potentially a cornerstone for success due to factors outside of our control influencing our emotions, as well as the potential for humans to change over time. This self-learning aspect is essential because the system will have to learn elements such as user preferences, which with the nature of humans, will change over time, and the system must be aware of these changes to react effectively to them.

One of the most critical aspects of affective computing, and the most critiqued, is its subjective nature. It is important to remember that the inner workings of the human brain, mind, and soul have not yet been unravelled. Hence the field is still developing, based on theories and assumptions until a firm psychological theory is developed on which to base the system. Therefore, it will also be impossible to predict the accuracy of the systems. The bulk of research, once a test system has been developed, will be for improving its accuracy through many of the technologies outlined here.

2.1.9 Emotion classification systems

Human emotion classification and recognition systems are still being developed within computing and other disciplines. One of the approaches suggested by (Mower, et al., 2011) is an audio-visual feature extraction, where the video is analysed to retrieve facial features and map these against emotional models.

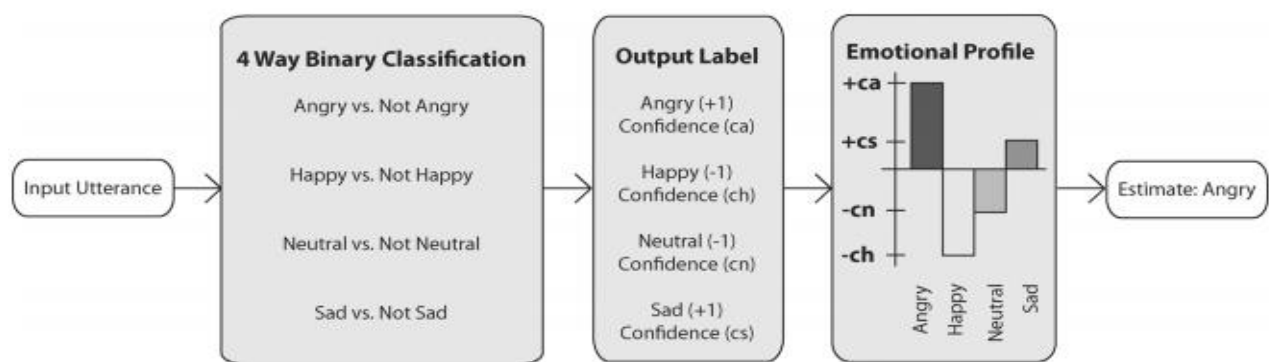


Figure 2:2 - Four-way binary classification (Mower, et al., 2011)

The model in Figure 2:2 shows one of the approaches used to interpret natural human expressions of combinations of underlying emotions. This model is designed to automate the processing of these emotions and reflect this processing in the form of a user emotion. One of the strengths of this method is its sensitivity to the selection of the base classifier. Base classifiers are probabilistic classifiers based on applying Bayes' theorem with strong independence assumptions between the features. Their benefits are that they are extremely scalable (Mower, et al., 2011).

The strength of this model is its straightforward approach, utilising binary classification for each emotion to provide a best-fit outcome. This system's methodology shows great

promise for a simple emotional recognition system and should work effectively with the methods outlined in this thesis.

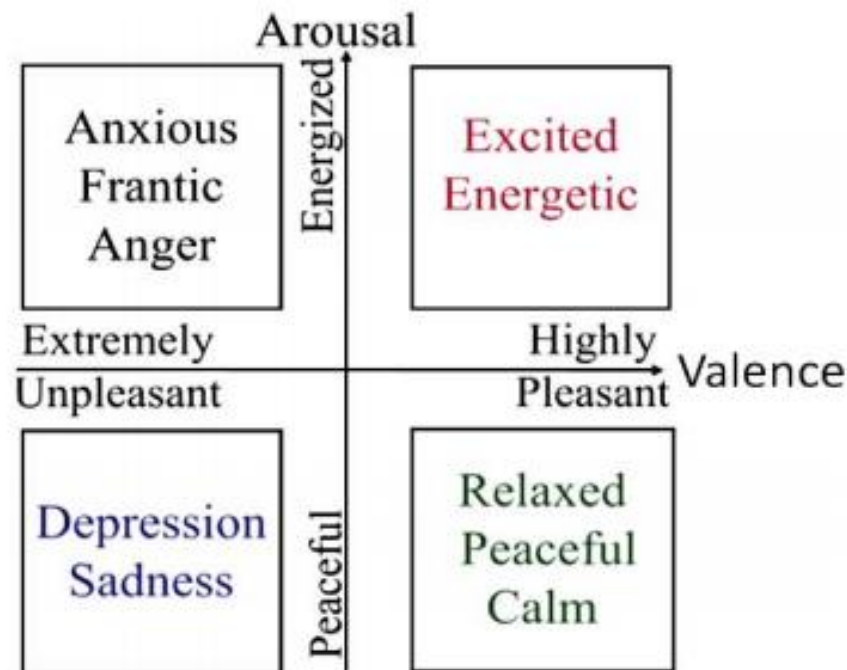


Figure 2:3 - Example of a dimensional affective model (Zhang, et al., 2010)

Figure 2:3 shows a popular psychological model in affective content analysis, the arousal—valence model deriving from the works of (Lang, 2014). These models' affective responses are represented using two components: arousal and valence.

This highlights the importance of the model in the work of (Niese, et al., 2011), which simplifies the emotional process to a reduced number of possible outcomes. With so many emotions, this research will require a scaled-down approach. However, this is not a classification system as such, but it shows the assortment of different emotional labels that could exist. Therefore, the first systems of this kind will need to be scaled back regarding the scope of emotions. In addition, the classification system will need to become more aware of

the development and develop a higher level of self-learning than possible within the scope of this project. The condensed emotional profile approach adopted in this study is shown in Figure 2:4.

2.1.10 Circumplex model

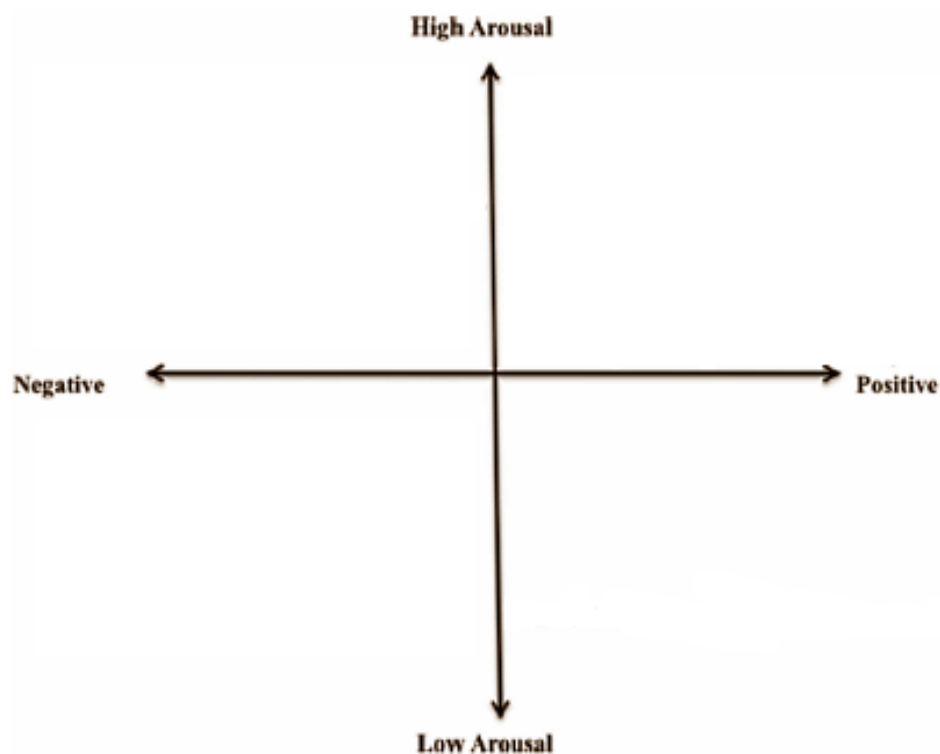


Figure 2:4 - Circumplex Model (Russell, 1980)

The circumplex model is a widely used emotional model developed by (Russell, 1980). The model assumes that emotion can be distributed across a two-dimensional grid containing arousal and valence dimensions. In Russell's model, the vertical axis represents arousal, and the horizontal axis represents valence. This model was based on the assumption that any emotional state can be represented at any point on two axes. The horizontal axis is considered a spatial metaphor and highlights pleasure–displeasure. The vertical axis can be described as a continuum from sleep to arousal (Russell, 1980).

One study that has expanded on Russell's earlier works was (Scherer, 2005), where 80 emotional terms were mapped onto the circumplex model shown in Figure 2:5. The research led to developments in the Geneva Emotion Wheel (GEW), covered in Section 2.1.11.

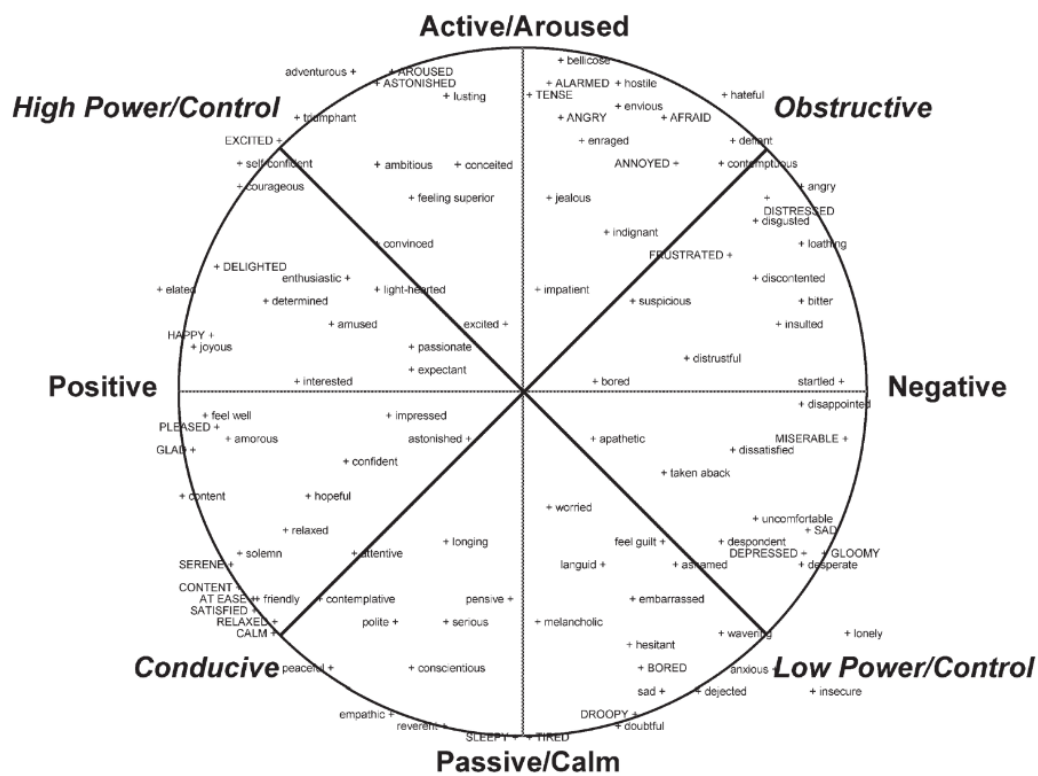


Figure 2:5 - Alternative dimensional structures of the semantic space for emotions (Scherer, 2005)

The circumplex model has been used in studies to categorise emotions. One such study used the circumplex model and simplified it through the discrimination of four categories, to which facial expressions were mapped, to provide one of four possible outputs (Niese et al., 2011) (see Figure 2:6). The circumplex model in Niese et al.'s work was condensed to describe anger, joy, boredom, and satisfaction. This simplifies the circumplex model as something users can use to provide direct feedback regarding perceived emotions.

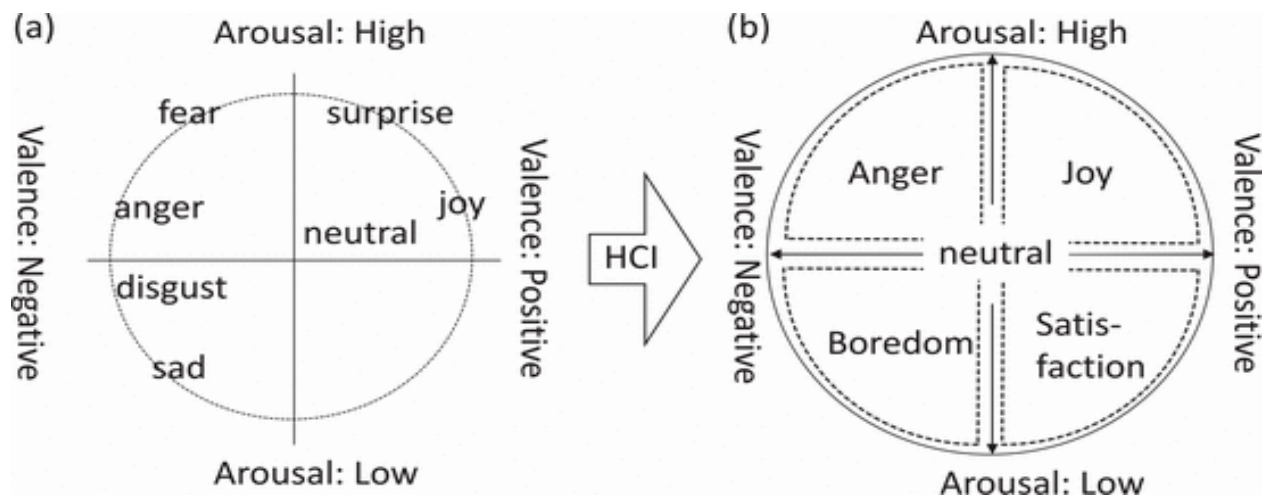


Figure 2:6 (a) Circumplex model of affect, (b) model simplification through discrimination of four categories (Niese et al., 2011)

Geometric facial features and their relation to facial feature points used to form facial expressions were analysed. The study's outcome was the ability to map facial expressions against the circumplex model to provide an emotional outcome. While this model was adopted for facial expression modelling, the simplified approach to emotional classification provides a testbed for the progression of the classification of emotional responses to films.

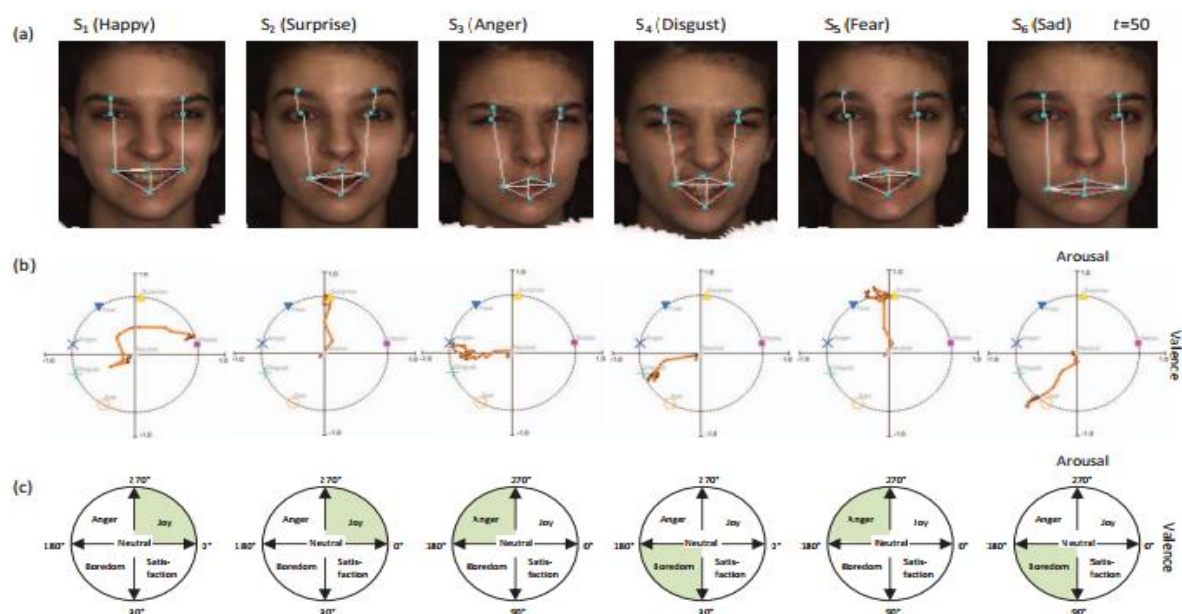


Figure 2:7 - Facial expression mapped to emotional outcome chart (Niese et al., 2011)

2.1.11 Geneva Emotion Wheel (GEW)

The Geneva emotion wheel is theoretically derived and empirically tested to measure emotional reactions to objects, events, and situations. Constructed of 20 distinct emotional families with a range on a wheel shape, its two axes are defined by arousal/dominance/control and valence of an emotional experience (Scherer, 2005).

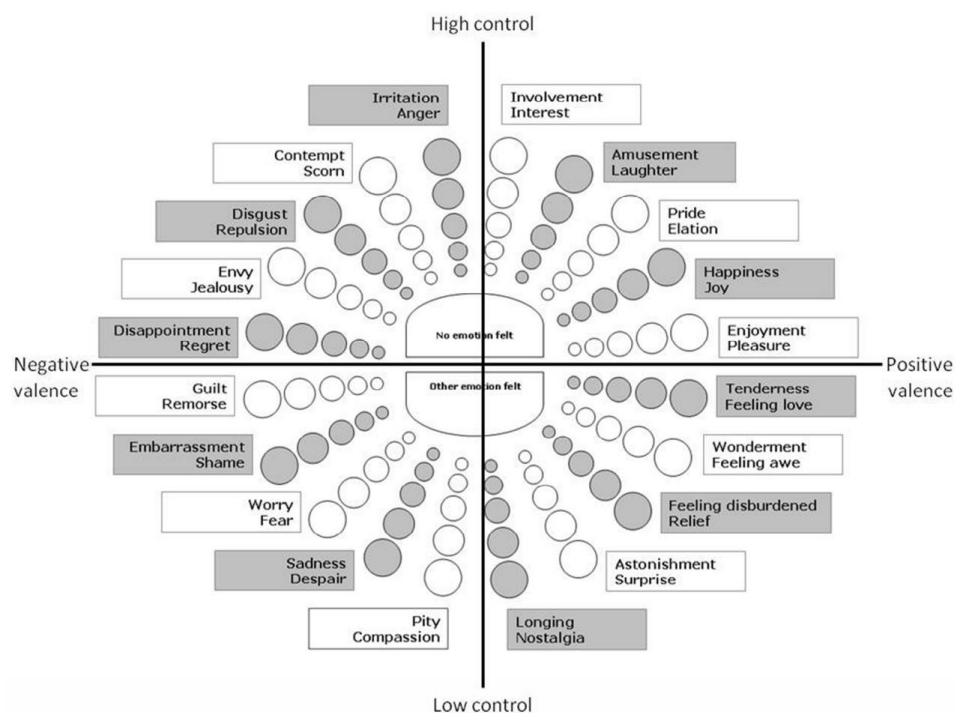


Figure 2:8 - Version 2.0 of the GEW with 40 emotion terms arranged in 20 emotion families (Sacharin, 2012)

2.1.12 Plutchik's model

Plutchik's model (see Figure 2:9) is a three-dimensional model of emotions; it arranges emotions in concentric circles. The inner layer illustrates the more basic emotional scales,

working out to more complex emotions on the outside. This model is called the wheel of emotions because it is a circular arrangement. It consists of eight primary emotional states, which are then broken into subcategories used for classification to determine the intensity of the emotion (Roeckelein, 2006).

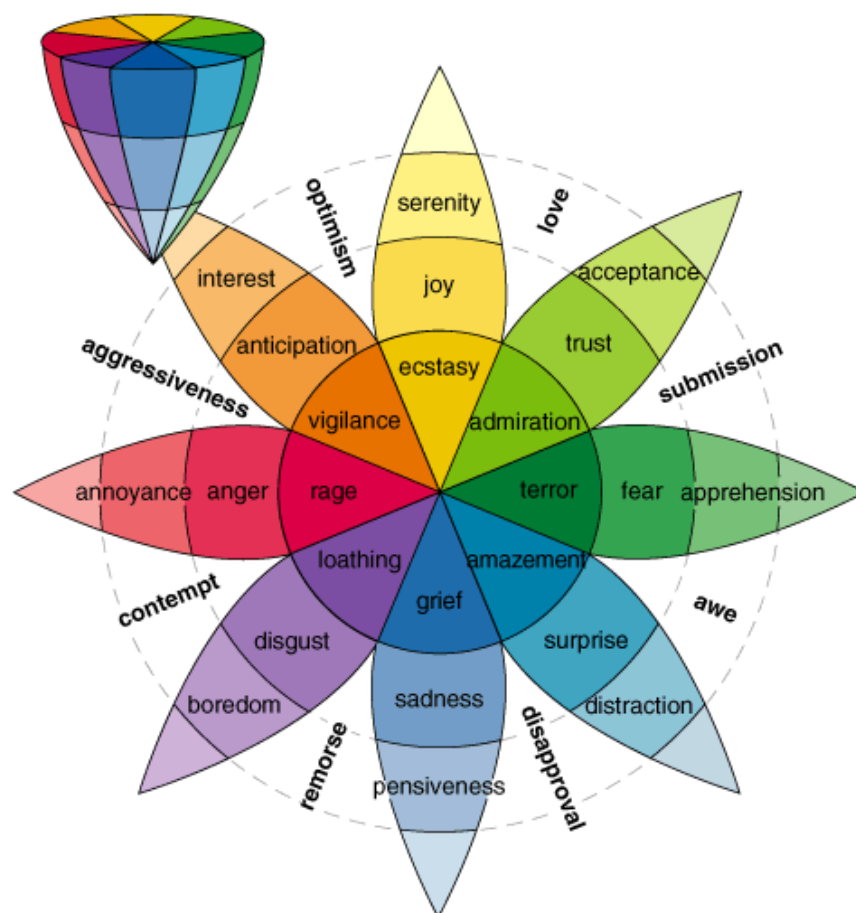


Figure 2:9 Plutchik's Model (Plutchik, 1997)

Plutchik adopted colour metaphors for his model to illustrate that basic emotions could be combined to yield a new emotion. Plutchik's theory suggests that each of the basic emotions has a distinct psychological basis and that these distinct emotions can be merged to form new emotions (Ortony & Turner, 1990).

2.1.13 Self-assessment manikin (SAM)

The SAM scale is used to evaluate valence, arousal, and dominance on one scale for affective reactions. It has become widely used in the affective field due to its simplicity and ease of use. The SAM scale expands on the three basic dimensions outlined in (Wundt, 1896), which were initially labelled lust, Spannung and Beruhigung (calm). The subsequent research confirmed that pleasure arousal and dominance are universally applicable when discussing human judgement for a range of perceptual and symbolic stimuli.

The SAM scale also addresses some of the issues found using the semantic differential scale, such as each rating: a 9-point scale resulted in requirements of the factorial analysis of 18 ratings and generated underscores of dimension pleasure, arousal, and dominance.

The other issue was its reliance on a verbal rating, making it difficult to use outside of a non-English-speaking culture. The SAM scale was developed to address these two main issues with previous approaches, using pictures and a shorter version of the scales (Mehrabian & Russell, 1974).

The SAM scale shown in Figure 2:10 is a graphic that ranges from smiling and happy to frowning and unhappy, representing the valence dimension. Likewise, for the arousal dimension, the SAM scale ranges from excited and wide-eyed to relaxed and sleepy. Finally, the dominance dimension represents the changes in control via the changes of the SAM, with a larger figure defining maximum control and a more petite figure defining minimum control (Mehrabian & Russell, 1974).

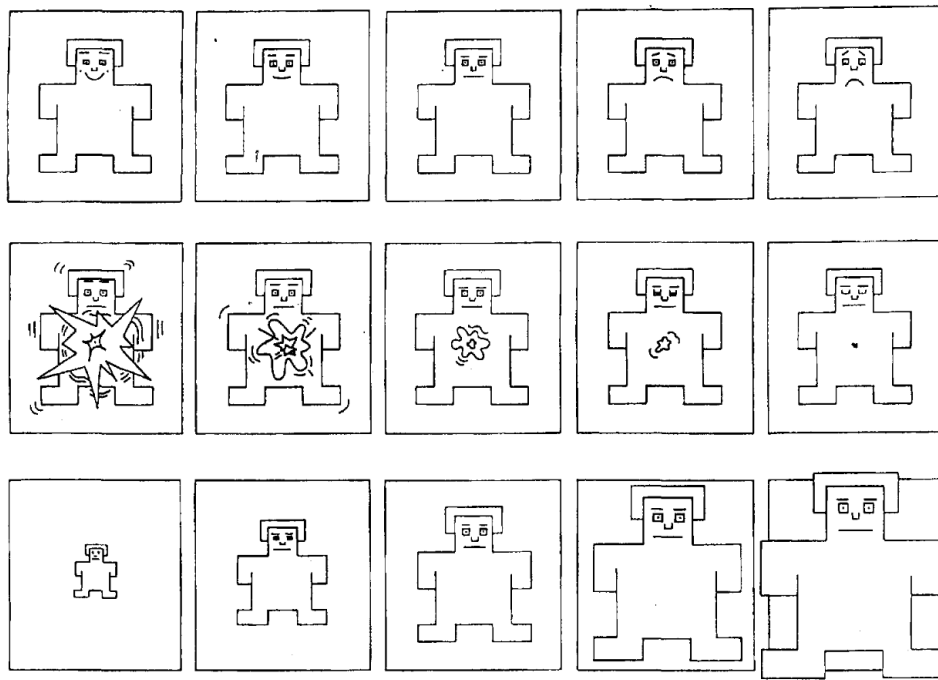


Figure 2:10 SAM scale dimensions of valence (top), arousal (middle), and dominance (bottom) (Mehrabian & Russell, 1974).

The SAM scale is versatile in its approach; however, the key is in its simplicity, as the user only needs to indicate where they believe they appear on the scale. In the original studies, this was primarily done with pencil and paper, as the users would mark an “X” on the aligned scale to provide their rating. Studies since have used radio buttons aligned directly with the SAM scale in some manner. More recently, research has developed the “affective slider,” which uses the philosophy of the SAM scale to create a digital self-assessment scale for measuring human emotions (Betella & Verschure, 2016).

2.1.14 IAPS

The IAPS was developed to provide a set of normative emotional stimuli for experimental investigations of emotion and attention. It comprises colour photographs that are standardised, emotionally evocative and internationally accessible, covering a wide range of

semantic categories. It was developed by the NIMH Centre for Emotion and Attention (CSEA) at the University of Florida to provide standardised materials for researchers studying emotion and attention (Lang, 1997).

IAPS is a collection of normatively rated affective stimuli and should provide the following benefits; Firstly, providing better experimental control in the selection of emotional stimuli. Secondly, to facilitate comparison of results across numerous studies conducted in the same, or different, experimental environments and certainly allow for exact replications across research labs when researching problems in psychological science.

IAPS relies on a simple dimensional view based on assumptions that emotions can be defined by a coincidence of values on a number of different strategic dimensions. This was based on views found in Osgood (1957). These are considered seminal works with the semantic differential, in which factor analyses were conducted on a wide variety of verbal judgements, which indicated that the variance in emotional assessments was accounted for by three major dimensions. The two primary dimensions were the affective valence, which ranges from pleasant to unpleasant, and the affective arousal which ranged from calm to excited.

A third, less strongly related dimension was variously called 'dominance' or 'control', which is not included in the AVS system. Dimensional views of emotion have been advocated by a large number of theorists throughout the years, including (Titchener, 1898), (Mehrabian & Russell, 1974) and (Tellegen, 1985). It is worth noting that dominance does not appear in the circumplex model, only arousal and valence, and it is not included in the majority of

studies that utilise the SAM scale to measure arousal and valence; therefore, it is not included as a parameter of this study.

The IAPS also employs the SAM scale, which is covered above in Section 2.1.13. The SAM scale was utilised to allow subjects to rate IAPS materials on dimensions of valence, arousal, and dominance. The SAM scale is considered a relatively easy method for quickly assessing valence, arousal, and dominance. In this part of the IAPS study, results were compared to a semantic differential scale. When the results indicated an extremely high correlation between the scores of valence and arousal (derived from the systematic differential rating) and those from the SAM scale, this suggested that the SAM scale is a quicker method to assess these fundamental dimensions of emotion.

The SAM scale data indicated that these ratings were constant when assessing between subject reliability all within subject reliability.

Citing an example with a mean rating of valence and arousal for these materials is highly internally consistent. The findings reported for a split-half coefficients for the valence and arousal dimensions were highly reliable ($p < 0.001$), for both pencil and paper ($r_s = .94$ and $.94$, respectively for 60 pictures) and computer administration formats of SAM ($r_s = .94$ and $.93$, respectively for 21 pictures). This indicates that the affective results remained consistent, even when subjects in different experiments rated the same pictures (Lang, 1997).

IAPS is based on a rating procedure for how these normative studies were conducted. In general, each picture set of 60 different IAPS pictures that was rated consistently varied in

valence and arousal. SAM scale ratings of valence, arousal, and dominance were made immediately after each picture.

2.1.15 LIRIS-ACCEDE

LIRIS-ACCEDE consists of diverse video content annotated on affective dimensions. It was designed to overcome the limitations of existing datasets and to foster research in affective video content analysis. The dataset consists of 9,800 video excerpts shared under the Creative Commons licence (Commons, 2021) to prevent copyright issues. In addition to the datasets provided, it also outlines an experimental protocol for ranking video clips along the induced valence axis. It utilised a crowdsourcing technique to gather 2-D valence and arousal ratings and is considered one of the larger affective video databases currently in existence that also includes annotations by a representative population.

LIRIS-ACCEDE is split into six “collections” Discrete LIRIS-ACCEDE, Continuous LIRIS-ACCEDE, MediaEval 2015 Affective Impact of Movies task, MediaEval 2016 Emotional Impact of Movies task, MediaEval 2017 Emotional Impact of Movies task and MediaEval 2018 Emotional Impact of Movies task.

The 9,800 excerpts that make up LIRIS-ACCEDE consist of 160 films covering nine representative movies genres, which include comedy, animation, action, adventure, thriller, documentary, romance, drama and horror, which was considered a normalised distribution of movie genres. This was based on data from IMDb (IMDb, 2021) and ScreenRush Data shown in Figure 2:11. It also tried to be representative of languages; while mainly English, there was a small set of French, German, Icelandic, Hindi, Italian, Norwegian, Spanish,

Swedish and Turkish films, subtitled in English, as well as 14 silent movies. The 9,800 segmented video clips last between 8 and 12 seconds, with a total running time of 26 hours, 57 minutes, and 8 seconds.

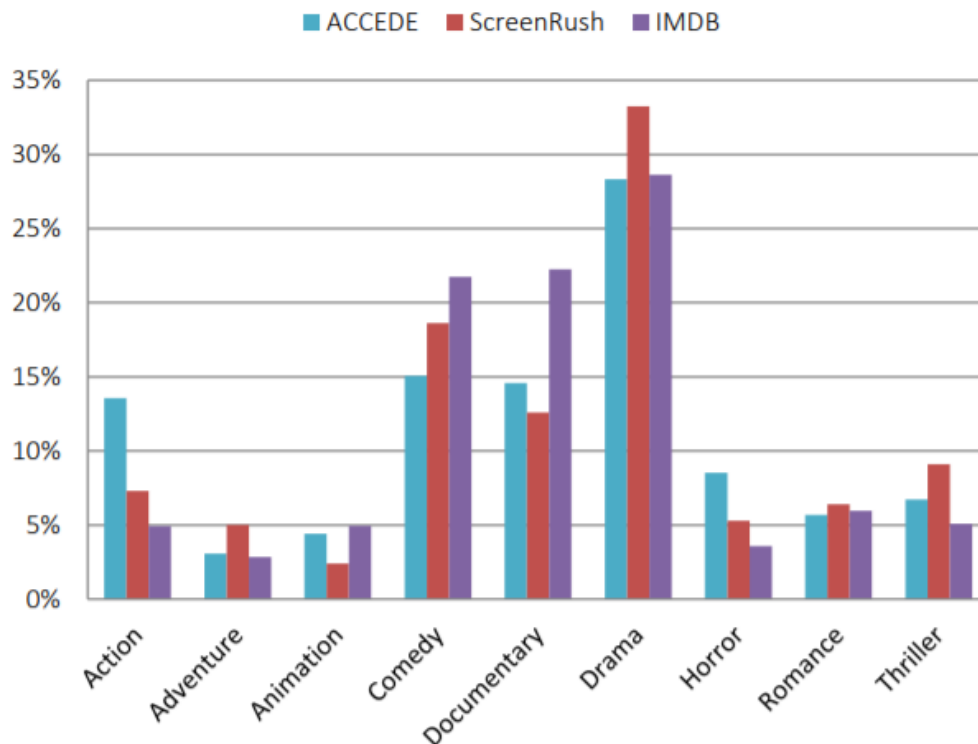


Figure 2:11 Normalised distribution of films by genre included in LIRIS-ACCEDE (Baveye, et al., 2015).

2.1.16 Film stim

Film stim was developed to assess the effectiveness of a methodological tool for emotion research. The research comprises two main parts: the first was how the film collection was created and the second was the effectiveness of films as emotional stimuli and how this was tested (Schaefer, et al., 2009).

The main focus of the study was to create a collection of emotionally-eliciting film stimuli that would be as flexible as possible to make it applicable across varying research domains.

It was not a goal of the study to investigate mechanisms of emotion, but instead, to develop an effective methodological tool for emotional research meeting the following goals:

- To select a large set of emotion-eliciting film scenes covering a wide range of emotional dimensions
- To include subsets of positive films that can elicit attachment-related emotions.
- To assess the effectiveness of the film set according to several criteria
- To provide a method of classification of films with multiple criteria and identify subgroups of films with high scores on every criterion
- To make the data freely available on the Internet allowing a flexible choice of stimuli by potential users (Schaefer, et al., 2009).

Their methods first looked into the selection of relatively large numbers of film excerpts corresponding to seven pre-existing emotions: anger, sadness, fear, disgust, amusement, tenderness, and neutral state.

2.1.16.1 Video content design

The way this study addressed film selection was to start with ten films based on emotional category, resulting in 70 film excerpts to be selected from individual film scenes. This stage was carried out with film rental store managers due to their expertise and knowledge of films. They were asked to fill in a short questionnaire to describe and provide an emotional rating for the excerpts for each emotional category. This resulted in a set of 824 film excerpts, from which the ten most frequently cited excerpts were chosen. Film clips and solved range from one to seven minutes, as each clip was cut into a logical segment.

2.1.16.2 Experimental design

The study involved 364 undergraduate students who were randomly assigned to one of seven groups. There was an equal proportion of female and male participants; each group consisted of an average of 52 participants.

The study was done in a laboratory in subgroups of three to five people. They were then briefed and instructed on what to expect. Participants were instructed to go through a relaxation procedure before viewing each excerpt, and they would complete a self-reporting questionnaire experiment after viewing the excerpt.

The instructions were based on the work of (Philippot, 1993), where participants were encouraged to first report what they actually felt and not what they believed people should feel in reaction to the movies, which strengthened the importance of individual differences. Also, they reported how they felt at the specific time they were watching the video excerpt and not the general mood of the day.

This procedure was used as a retrospective evaluation of emotions, as it was known to be a good predictor for the actual state felt during the excerpts (Fredrickson, 1993).

2.1.16.3 Experimental measures

The measure of the experiment was self-reported emotional arousal, where participants were encouraged to answer the questions in accordance with what they actually felt during the test. The intensity of subjective emotional arousal was assessed using a 7-point scale:

“While I was watching the film ...” (1)“I felt no emotions at all” to (7)“I felt very intense emotions”.

The differential emotional scale (DES) is utilised to assess discrete emotions because it is one of the most widely used self-report scales for discrete emotions. The study also made some modifications to DES selected add additional items and had been previously tested in other studies (Schaefer, et al., 2009).

PANAS was used for self-reporting adjective checklists for the assessment of positive affect, which rated the extent to which they felt each state while watching the excerpt.

2.1.17 The emotional movie database (EMDB)

EMDB was developed as a study to develop a new database of affective film clips without auditory content based on dimensional approaches to emotional stimuli of valence, arousal, and dominance (Carvalho, et al., 2012).

The study was comprised of three different phases:

- Phase 1 was the preselection and editing of 52 film clips
- Phase 2 was a self-reporting of these film clips by 113 participants
- Phase 3 psychophysiological assessments of skin conductance levels (SCL) and heart rate in 32 of the volunteers.

The research sought to expand on the IAPS, which has had a notable impact on the study of emotions and attention. Noted limitations were the utility of this emotional database, namely habituation-related reactions of repeatedly presenting the same picture to the same subject (Dan-Glauser, 2011). Another problematic limitation is the attenuation of the

emotional impact of the stimulus as time increases, as highlighted in (Koukounas, 2000), which has led to other researchers using blocks of several images (Gomez, 2010).

The present research suggests that using emotionally arousing film clips instead of pictures may help to overcome these limitations, referencing a study where emotional elicitation using films was conducted (Rottenberg, 2007).

2.2 Recommendation systems

Recommender systems try to find 'common interest' between elements of data to provide relevant outcomes, for example, where data on user movie buying preferences are collected and stored. A later recommendation for which movie to buy is then based on user interests collected from historical data on the individual's movie buying behaviour to provide relevant outcomes (Jannach, et al., 2012).

Recommendation systems require the integration of affective data to support intelligent decision-making processes. Vast amounts of other data are also relevant to recommendation systems and are available in datasets. This data helps to improve decision making, together with affective data.

One of the adopted mechanisms for recommender systems is collaborative filtering (CF). This does not exploit or require any knowledge about a specific user, as it accesses other users' data and makes recommendations based on that knowledge base. Whereas a knowledge-based recommendation system requires a historical data set within its area for a specific user to make a decision, CF aggregates data from multiple users (as explained in the following section 2.2.1). However, knowledge-based systems have an inherent problem for

a first user, as there is no historical data upon which to base a decision. Hybrid approaches, therefore, try to counter this problem (Jannach, et al., 2012) Hybrid approaches are discussed in Section 2.2.6.

One of the challenges for recommender systems is the mass of data and how to break down some of the decisions in such a way that the system makes a relevant decision most of the time, using the many factors that may influence our everyday lives. With the emergence of streaming services such as Netflix and Amazon Prime, users now have access to endless lists of films and book titles, further highlighting the need for such systems to support user decision making due to the sheer volume of data.

2.2.1 Collaborative filtering (CF)

Collaborative filtering (CF) is based on an algorithm that outputs the ratings or behaviours of other users within a given system. The main assumption of this type of recommender system is that users' opinions can be aggregated in such a way as to provide an informed prediction. Data are collected on users' perceptions of the quality of an item. User preferences are mapped and compared with other users' perceptions of quality to find common interests among users so that decisions can be made based on collective behaviour.

CF techniques depend on several concepts to describe the problem of the information domain. CF systems consist of *users* who will have expressed a preference for various *items*. Once the user has shown a preference for an item, this becomes known as a *rating* and is

represented as a (User, Item, Rating) triple. Once the required elements of the triple are gathered, they can be used to form a sparse matrix commonly referred to as a *ratings matrix*. Where the user has not expressed a preference for a ‘rated’ item, these are unknown values to the matrix (Ekstrand, et al., 2011).

2.2.2 Baseline predictors

$$b_{u,i} = \mu$$

Equation 1 - Simplest Baseline (b = baseline rating; u = user; i = item; μ is the overall average rating)

Baseline prediction methods are useful for establishing non-personalised baselines against which personalised algorithms can be compared. Baseline predictors correspondingly provide grounding for pre-processing and the normalisation of data for use with more sophisticated algorithms. Simple baselines can predict the average overall rating in the system (Ekstrand, et al., 2011). Therefore, for a user (u), the baseline (b) for an item (i) is defined as the average overall rating available (μ).

2.2.3 User–user collaborative filtering (CF)

User–user CF was one of the first automated CF methods. The algorithm represents the core principle of CF, which is to find other users with a past rating behaviour similar to the current user and employ their ratings on other items to predict what the user would like. However, before generating a prediction/recommendation for a given user, a *neighbourhood of neighbours* must be formed; this is used to group users into similarity groups on which to base the prediction and recommendations.

The following scenario is based on selecting a television programme or film for a group of users who have watched shows One and Two and then proceed to watch show Three. The recommendation for show Three may be presented to users who watched Shows One and/or two. Users who have watched Shows One and Two may therefore also watch Show Three, which is offered to them. In this way, the item and the user share a direct relationship, with the user influencing the item and the item influencing the user's decision.

One of the elements of user–user CF is the similarity function, for which there are several statistical calculations available:

- **Pearson correlation** (Pearson's r): The Pearson correlation is a method that computes the statistical correlation between two user ratings of the same item to determine how well they are related.

r = correlation coefficient

i = index

x_i = values of the i -th x -variable in a sample

\bar{x} = mean of the x -variable

y_i = values of the i -th y -variable in a sample

\bar{y} = mean of the y -variable

$$r = \frac{\sum (x_i - \bar{x}) (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Equation 2 - Pearson's correlation coefficient formula (statisticshowto, 2022)

2.2.4 Content-based filtering

Content-based recommendations include classification tasks from a machine learning perspective. These classification tasks, based on content, contribute to the decision-making process. A specific classification rule is created for each user, which is based upon user rating information and the attributes of each item (content), leading to the categorisation of whether an item would be interesting to the user or not.

Although a sophisticated technique, content-based filtering is less successful than other methods due to the scarcity problem, which is loosely defined as unlimited human wants in a world of limited resources (Felfernig & Burke, 2008). An individual user may not provide sufficient data for a reliable profile to be generated. Techniques that have proven reliable in content-based recommendation are based on computations of k-nearest neighbours and naive Bayes (Felfernig & Burke, 2008).

2.2.5 Knowledge-based recommender system

Knowledge-based systems rely on knowledge sources that are commonly associated with CF or content-based filtering. These systems focus on the user's situation and on how recommended items can fulfil a particular need for the user.

They typically have a much higher level of complexity than their counterparts due to their reliance on historical information. The two well-known approaches are case-based recommendations and constraint-based recommendations.

2.2.5.1 Case-based recommendations

The case-based recommendation treats recommendations as similarity assessment problems, where the system finds an item that is most similar to what the user was searching for.

2.2.5.2 Constraint-based recommendations

The constraint-based recommendation works on the basis of explicitly defining constraints; if no item fits the criteria of the user's constraint base, a similarity value is calculated to provide the result (Felfernig & Burke, 2008).

2.2.6 Hybrid recommender systems

Hybrid recommender systems combine different approaches to improve effectiveness. Hybrid approaches have emerged to overcome the known problems of other systems (as identified in the following section); they combine multiple techniques to achieve synergy and address these problems.

2.2.7 Summary of recommender systems

Recommender systems play a particularly key role in accessing information for multimedia decisions. For example, a method such as CF requires an array of data to be analysed. If a user, watches Film 1 and likes it (rating it highly), and watches Film 2, (rating it highly), and a second user watches Film 1 (rating it highly), they will be recommending Film 2.

Content-based approaches use system-generated recommendations from two sources:

- Features associated with items and the ratings that a user has given them

- Content-based recommenders treat recommendations as a user-specific classification problem and learn a classifier for the user's likes and dislikes based on item features (Felfernig & Burke, 2008).

A knowledge-based recommender system suggests items based on suggestions about a user's needs and preferences. This knowledge will sometimes contain explicit functional knowledge about how certain product features meet user needs; this knowledge may be gathered from external sources as well as from users.

Although considered effective, one of the drawbacks of the user–user CF system is its lack of scalability. As the user base grows, the search for neighbours increases proportionally. This has led to the emergence of item–item CF, which has become widely adopted. This uses the rating patterns between items as opposed to the rating patterns between users and has overcome the scalability issue.

Content-based recommender systems have access to item features, such as keyword categories, and therefore do not suffer from the 'new item problem'. However, they do suffer from the 'new user problem' so users must build up profiles with multiple ratings.

The quality of the data collected is of primary importance in content-based systems. The profiles generated for each user can only be as good as the system's data collection and the relevance of the data in representing distinctions in the domain (Felfernig & Burke, 2008).

2.2.8 Machine learning

Machine learning is a set of methods computers use when making predictions or behaviours based on data. The strength of machine learning methods is that the machines surpassed

humans in many tasks, such as playing chess or, more recently, computer games. Its second most advantageous quality is speed while being able to scale and be repeatable. Ultimately, once implemented, machine learning models can complete tasks much faster than humans as well as deliver reliable and consistent results (Molnar, 2020).

When utilising machine learning methods, the following steps must be followed. Step one consists of data collection and machine learning, the more the better as long as the collected data is related to the desired outcome. Step two is entering this information into a machine learning algorithm, which then creates a model based on the data provided. Step three consists of the utilisation of the constructive model against the new dataset (Molnar, 2020).

2.2.9 Regression

When investigating relationships between variables, regression analysis statistical techniques can be utilised (Montgomery, et al., 2013). The combination of machine learning using regression forms a strong platform for testing forthcoming datasets containing arousal and valence values collected from users against independent variables from video content.

2.3 Related work

This section reviews related work in the areas of affective computing, context awareness, recommender systems, and context-aware emotional recognition services. These systems have informed the later chapters.

2.3.1 Emotional context awareness

A study conducted by (Yoo, et al., 2011) looked at the role of emotion within context-awareness services, highlighting that context-awareness systems must provide the user with a natural interface, so the user feels comfortable with the services. Their study demonstrated that users' emotions are sometimes observable and that these observable emotions affect users' decision-making behaviours. Estimating or predicting a user's emotional state was therefore concluded by (Yoo, et al., 2011) to be crucial in higher-level context decisions to improve decision quality.

The paper proposes an emotional estimation methodology for context-aware services. The proposed method from this paper will look to process the real-time emotional states of the user using a new methodology and recognises two-dimensional emotions that users experience day to day. This method uses audio analysis of spoken words in a ubiquitous smart space environment, making use of Russell's circumplex for perspectives of emotions. Words were mapped against the circumplex model in the order of the emotional states they represented (see Figure 2:12).

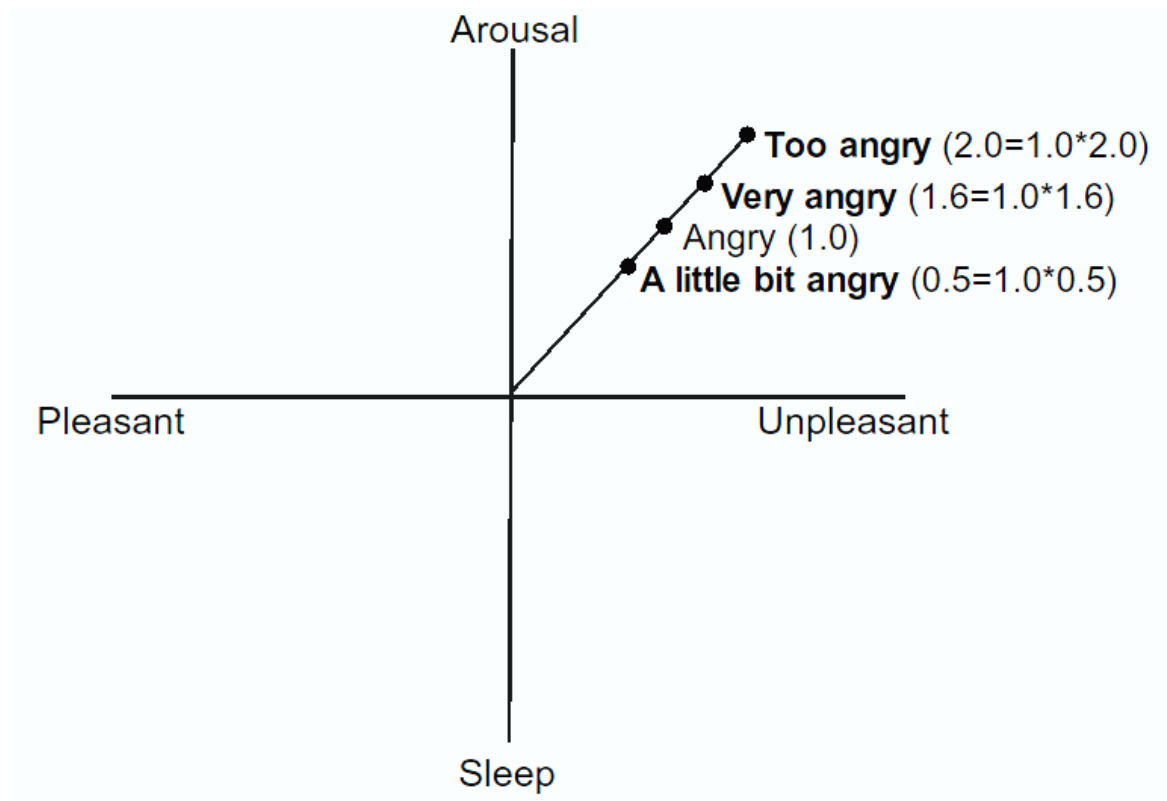


Figure 2:12 - Example of the degree of emotional scaling (Yoo, et al., 2011)

Figure 2:12 outlines the method used for emotional analysis; it also shows the methods used to calculate the mathematical value for the emotion of words onto the circumplex model. This work then proceeded to introduce “*mixed emotions*,” showing how the model would address multiple emotional inputs.

The research concluded by stating that emotion plays a key role in terms of providing a personalised context-aware service. One weakness of this study is that “spoken” words may lead to inaccurate results; humans may not verbalise or disclose their true feelings.

However, the methods outlined, when applied with audio analysis of stress levels in voice, for example, could yield more accurate results, due to this physiological element being hard to misrepresent, being based on involuntary or subconscious stimulations (Yoo, et al., 2011).

2.3.2 Affective methods for TV services

The following study investigated gathering the affective state of an audience while watching a movie based on the physiological responses of a real audience. Some typical signals are electrodermal activity (EDA), heart rate variability (HRV), or facial surface electromyogram (EMG), while evaluating the value of these physiological signals, with a view of providing an affective scoring for movies (Fleureau, 2013).

The experiment was conducted on a real audience during regular cinema viewing. Two films were shown to the participants. An EDA sensor was placed on the palm area of each participant. A questionnaire was used to report on the different scenes in the movie and their relevance in terms of emotional impact.

The data analysis method was done with box plots used for every individual affective profile (IAP) as a representation of the distribution of the IAP for every user at a given time, where the central mark was the median and the edges of the box were the 25th and 75th percentiles.

One distinct strength of this study was its ability to analyse the temporal and dynamic aspects of arousal during each movie's timeline shown in Figure 2:1. This provided an excellent insight into flashpoints of arousal while users were viewing it. Another advantage of this study is its use of physiological biosensors to make the process less intrusive, which is an important aspect for the future of this technology.

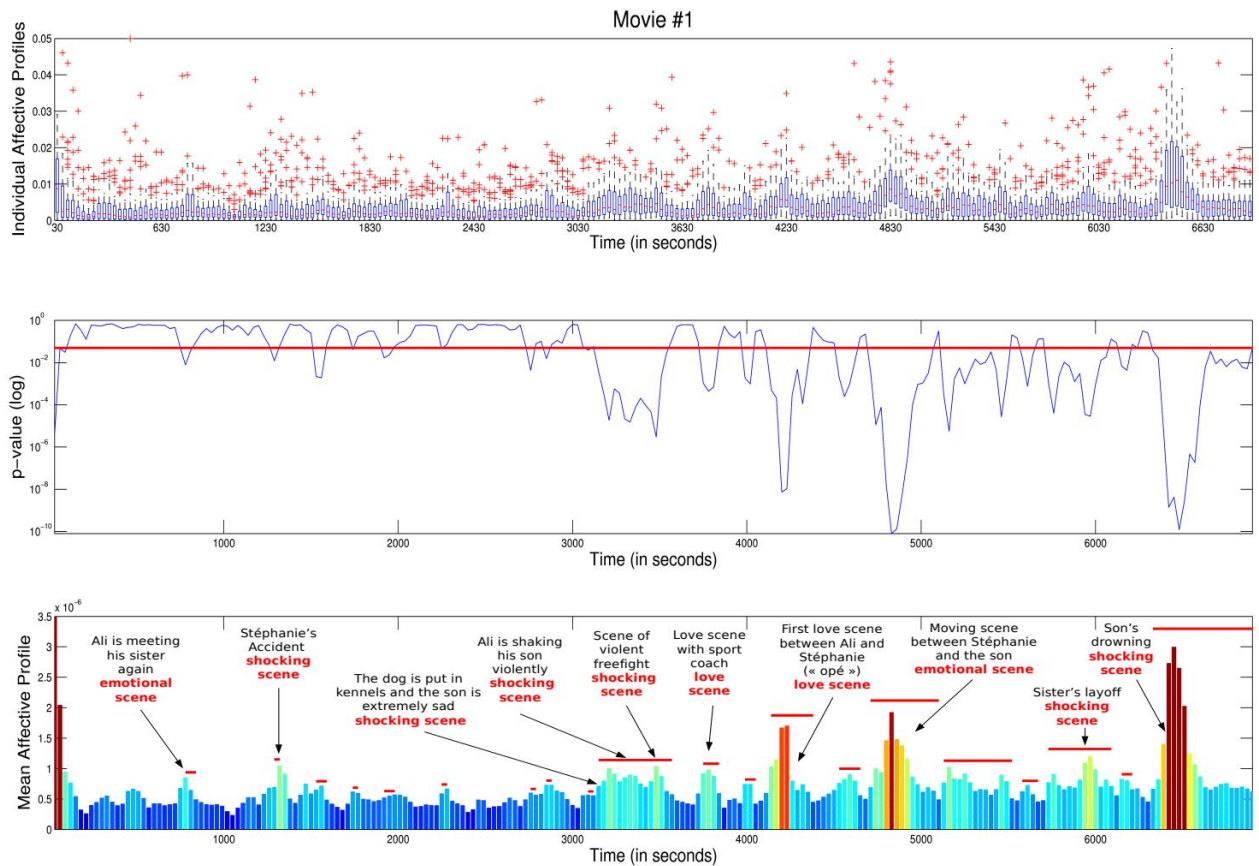


Figure 2:13 - Affective benchmarking solution applied to movies (Fleureau, et al., 2013).

Currently, it is hard to discern and define the sensory meaning of these sensor outputs, and the analysis is largely based on the analysis from the timeline of the event in the video content. Also, there is no accounting for pre-existing personal preferences of the users included within this study.

2.3.3 An evaluation of the International Affective Picture System

This paper highlighted the personality and individual differences between users. The aim of the study was to attain normative ratings of selected pictures from a population consisting of elite college student athletes. The purpose of the study was to provide normative ratings that may be useful for researchers studying emotions under laboratory conditions.

The study consisted of 219 participants (59 females and 160 males). These participants were considered elite athletes. However, while the sample size is respectable, the ratio of females to males is not, meaning it is not a fair representation. This method could not be used for average users due to the elitist nature of the study.

One of the strengths of this study were that it had a large sample size of 219 participants. However, the study was not representational or diverse because of the 'elite college students' (Tok, 2010) that were used to conduct the research.

The Individual Zone of Optimal Functioning (IZOF) model scale used was well-founded for this study due to the athlete population the study was conducted with, but Ekman's scale is generally considered the best in this field for this type of study. The strength of this study lies in its appropriate use of the circumplex model, and the method used to create "one value" is interesting.

The data analysis method was the five-factor personality inventory, which was used to assess main personality traits. A Pearson product-moment correlation between Turkish and American norms regarding valence and arousal dimensions for selected pictures was calculated. The locations were determined on valence and arousal dimensions, and the mean valence and arousal scores of each picture were calculated. A scatter plot, with the x-axis indicating valence and the y-axis arousal, was drawn in accordance with these two scores.

Each picture was represented as one score instead of two combining arousal and valence.

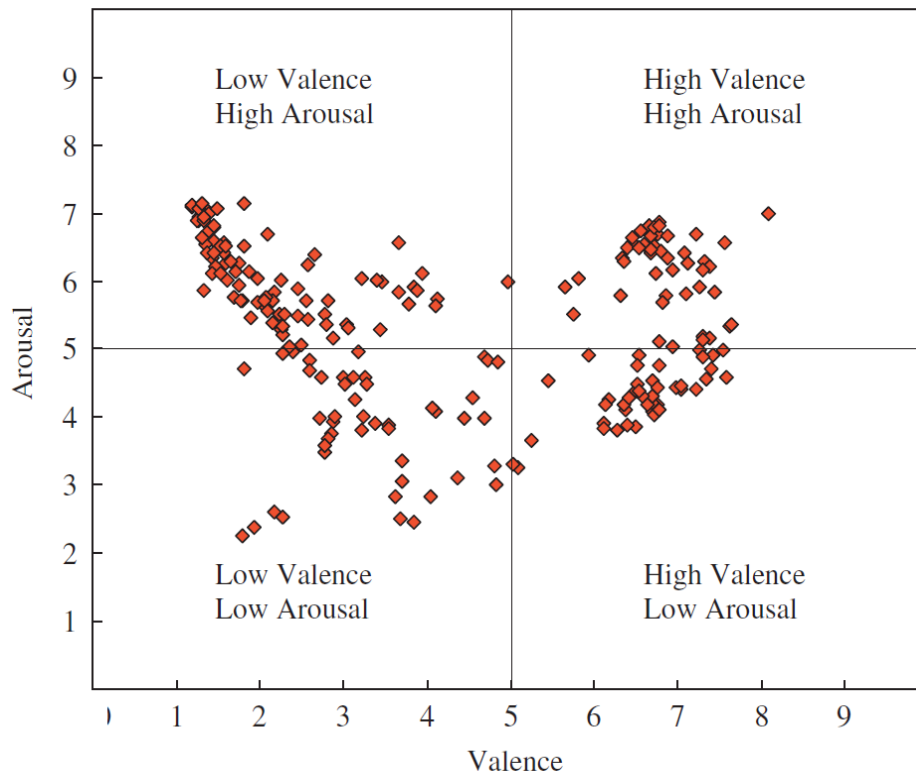


Figure 2:14 - Scatter diagram of pictures by arousal and valence dimensions (Tok, 2010).

2.3.4 iFelt: Accessing Movies through Our Emotions

iFelt is a study that developed a web application that allows users to catalogue, access, browse, and visualise movies and movie scenes based on their emotional properties. It was developed to explore the affective dimensions of movies. They presented a methodology which utilised the emotional labels; happy, sadness, surprise, fear, anger, and disgust for its classification of contents and users perceived emotions (Oliveira, et al., 2011). The reasoning behind this study was to address some limitations regarding discerning emotions from physiological signals, and that Ekman's basic categorical emotions are the most agreed upon.

2.3.5 Mining affective context in short films for emotion-aware recommendation

Paper presented by Orellana-Rodriguez, et al. (2015) explores two approaches to short film emotional association. Annotations were obtained by crowdsourcing on Amazon Mechanical Turk, and then a subsequent study was conducted to better understand the affect of those extracted by human intelligence. The other method was the automatic extraction of affective context from user comments available on YouTube. These were evaluated for the usefulness using an emotion-aware recommendation task.

2.4 Summary

Following the literature review's outcome, several areas of interest and gaps will require further study. One of the areas of significance is Russel's Circumplex Model, covered in section 2.1.10, which came up in several studies of interest. The study (Yazdani, et al., 2013) examined multimedia content and analysed it for the emotional characterisation of music video clips. This research discussed multimedia content and how to represent it at the affective level. It outlines the most common basic emotions of fear, anger, sadness, joy, disgust, and surprise, which were first introduced by (Ekman, 1999). The paper then explained that with a dimensional approach to emotions, the circumplex model represents emotions on dimensional space as coordinates. Highlighting that the goal of the dimensional approach is not to find a finite set of emotions as with a categorical approach (Yazdani, et al., 2013).

Moreover, in psycho-physiological studies with experimentation, it has been revealed that only certain areas of three-dimensional space are relevant. That valence and arousal were adopted into content analysis studies as a simplified model citing studies by (Xu, 2005) and

(Sander Koelstra, 2012). These studies (Annika Waern, 2009) and (Yazdani, et al., 2013) also adopt the 2D circumplex model for their studies. This led to the decision to move forward with the 2D approach for the research dropping the dominance scale as this would have needed a 3D model. Therefore, the circumplex model was utilised in the above studies and forgone the dominance scale.

This approach means that the IAPS and Circumplex Model can now be utilised to provide a firm footing for this research moving forward. As tools to build a firm test bed of research, these well-adopted approaches work together, and there is a precedent from previous works.

Chapter 3 - Affective Video System Framework (AVSF)

This chapter synthesises previous literature research in Chapter 2 and builds on it to propose a research approach for an affective video system. The research approach encompasses the theory behind affective methods, collecting and interpreting affect, and how the collected affective data can improve current content categorisation methods and recommendations. In addition, relevant research and technologies, including affective computing, IAPS, and machine learning, must be combined effectively to achieve an affective video system.

The study used the well-recognised IAPS (Lang, 1997) for the groundwork of the research and developed this into a transferable method to provide affective scoring for video content. This approach adopts the SAM scale (Bradley & Lang, 1994) as it aligns with the affective rating scale within the IAPS dataset and methods. This system scoring is then developed and used to find the best machine learning model by looking for significance within the dataset.

3.1 AVSF Overview

The AVSF framework was developed to support the development of any Affective video application. The framework's strength lies in being broken down into inputs, processes, and outputs, making it a higher-level framework with little to no restrictions on application.

This framework is based on breaking Affective methods into three main sections to map a framework that can be applied to this research problem. The rationale for the generic model is that as the system progresses and evolves, there will likely be a recurring trend of

cross-fertilisation between various datasets to provide affective processes and outputs. This means the methods must apply to any situation, not just specific situations. The benefit of this type of approach is that it does not require the development of numerous approaches to affective video systems.

3.2 AVSF Framework Overview



Figure 21 – Affective Video System Framework

Definitions	Description
Inputs: This is considered any data that can be derived from any data source that bears direct relevance to the problem it is being applied to in this example affective video. For instance, in this research, this could relate to a video feature.	This layer is concerned only with physical data that can be passed on to the process layer. It is loosely based on the same principles as a heterogeneous network; it may interconnect with various sensors, sensor networks, and data sources. This approach works well with the predominance of the Internet of things (IoT).
Process: This is how the data provided by the input layer can be analysed to provide the desired outcomes. For example, this research uses video features and regression methods to process these features.	This layer is where most of the complexity of the affective methods is broken down, drawing on multi-disciplinary techniques. Other elements in this layer include the specific user profile and mechanism linked to the self-learning aspect for any future development of the AVS system.
Outputs: These are categorised into two distinct outputs. Generic outputs are outputs that can be put out to a subset of users, and secondly, personal outputs linking directly to	Generic affective outputs are based on low-level decisions that could apply to multiple users. Finally, personal outputs represent a high-level affective decision that would need further research to

what the user would like to see as their desired outcome.	explore, but are required for the completeness of the model.
---	--

Table 3:1 AVSF Input, Process and Output approach

3.3 Affective Inputs

As a starting point, the AVS scoring system for video content is the primary source of decision-making. However, there is much more scope for additional input as the research evolves. Data inputs need to have an enormous scope, because the more accessible the information is at this layer, the more informed the decision in the next layer will be.

The main elements that form this layer are as follows:

User: User data can be collected from various sources. It could be used with affective keywords from AVS scoring to present recommendations to the user/s.

Environment: This data was collected from elements considered to be within the environment that can provide, for example, situational awareness. Smart devices can provide a wealth of information for this data source.

Middleware (AVS Scoring System): The middleware component acts as the communicator between different systems, consolidating into one system. The requirement for middleware is that it can eventually draw upon many varying data sources to allow improvements to its decision-making processes. This means that the middleware element becomes fundamental to allow this communication to happen, expand, and grow the system. Its addition is to help address scalability issues that these types of systems may encounter in the future.

3.4 Affective Processing

This layer forms the affective computing processing stage of the project, with the processes focusing on the methods to sort the presented data from the input layer. Below is a basic example of two essential parts of the AVS system.

User Profile and Modifications: The user profile stores personal information on the user to improve the query methods by saving successful and unsuccessful decisions.

AVS Database: This database holds information from the AVS scoring data that is crowdsourced against films to provide affective scoring.

3.5 Affective Aware Outputs

This layer forms the affective computing outputs stage of the project, with the outputs focusing on end-users. The two categories of outputs are as follows:

Personal Outputs: Personal decisions need to be based on modifying a decision to meet stakeholders' personal tastes.

Affective Outputs: The affective classification of outputs is used where an affective decision is required for presenting required video content to users.

Figure 3:1 synthesis of a research approach adopted to guide this research into the AVSF framework. The next stage is system design and studies to provide affective scoring of video content.

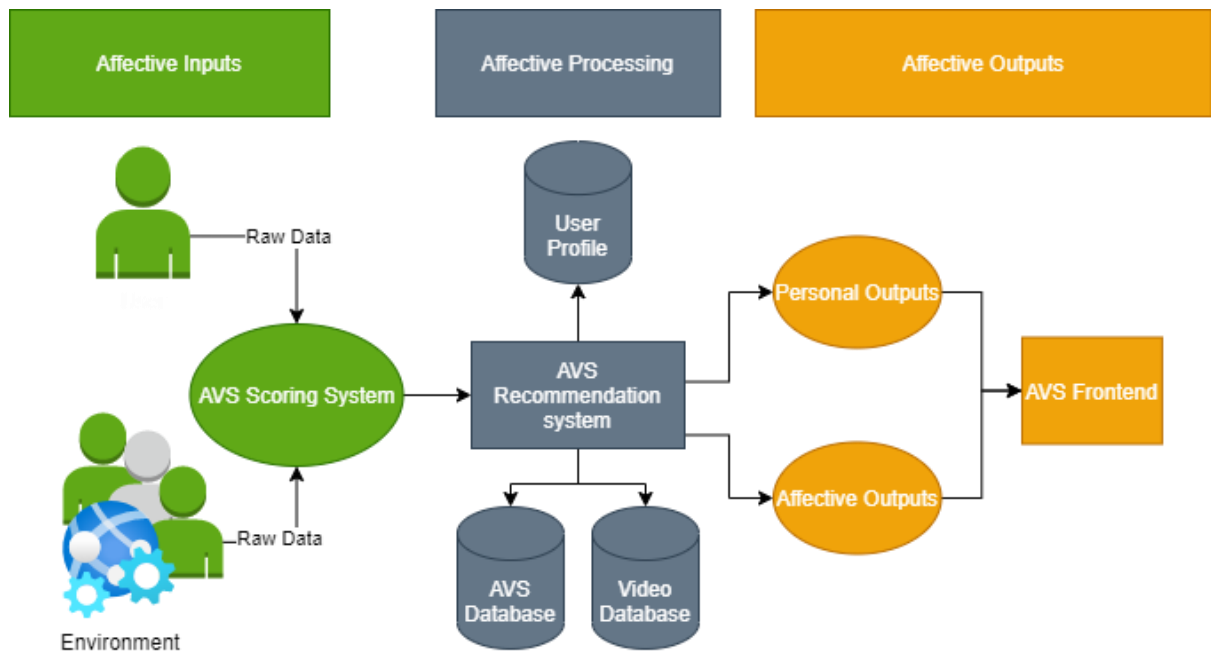


Figure 3:1 AVS Framework

3.6 Conclusion

This chapter has provided an overarching framework based on problems highlighted in section 1.1, where the objectives for affective computing can be positioned, allowing for building the following stages of experimentation. The model is based on well-established input, process, and output models.

Chapter 4 - An International Affective Picture System (IAPS) Approach for Affective Video System (AVS) A Rating Methodology for Affective Video

4.1 Introduction

This study's purpose was to look for comparable affecting scoring to similar video content to its original IAPS counterpart. This would be the score for the corresponding image in the IAPS data, as the AVS used video content that has been mapped against the original IAPS images. The video content was of a similar nature to pictures within the IAPS, allowing this approach to be used as a foundation for benchmarking affective video content against an internationally recognisable approach that has already been utilised in this field of study, but not directly for video content.

This study replicated and extended the international affective picture system principles into a method that can attempt to benchmark video content using internationally recognised methods within IAPS. The IAPS approach is a recognised benchmark for images widely used in affective fields and has been cited in over 3,700 studies.

The experiment considered a two-dimensional valence and arousal model based on scores provided via the self-assessment manikin (SAM) for reasons outlined in section 2.4. The scores from the experiment were compared with the original results provided in the IAPS. The data analysis attempted to determine whether there was a significant relationship between the IAPS images and the video content. The hypothesis was that the ratings for AVS video versions of the corresponding IAPS pictures should be comparable to the ratings for the pictures themselves.

This research aimed to move from picture-based IAPS to a video-based system, as video content is becoming increasingly popular with speed improvements on broadband networks. Cisco reported that video would be 82% of all IP traffic worldwide in 2021 (Cisco, 2020). The importance of this research is to move a well-rounded methodology into the future, and provide the foundation research for it to be used on the same basis as the original IAPS for video content.

The system, in this sense, is used in the same context as it is within IAPS. The importance of this type of system is due to the ever-increasing access to video content via the internet. The inception of this research consideration was primarily around video on demand. However, this has been superseded to a great extent by the emergence of user-created content, in which platforms like YouTube are created and endless sources of data to be processed. Furthermore, increasing the pressure on current recommendation methods far beyond their intended means, crowdsourcing is still at the heart of most systems. However, it has been simplified on many platforms over the last few years.

Due to the nature of categories and the way they are utilised on platforms like Netflix, the genre-based nature of this approach could supplement and expand this current area by using pressure on systems like recommendation systems, because they would be additional categories outside of a genre-based system.

Another challenging prospect within this research area is the subjective nature in which these decisions are made by users and considered highly personable. This means that as this

technology grows and evolves, this will be at the heart of its success; however, this is not within the scope of this study, which focuses on presenting the initial building blocks based on pre-existing methods to ensure a unified approach to affective video content. This is one of the most significant issues at this time, in my opinion, as there are so many varying approaches that it is impossible to adopt just one, and compare the merits of research directly. Hopefully, this research at least allows for a step in the right direction, based on a highly successful affective system developed for images.

The International Affective Picture System (IAPS) was developed by NIMH Center for Emotion and Attention (CSEA) to provide a set of normative emotional stimuli for experimental investigations of emotion and attention. Furthermore, the IAPS is presently used in experimental studies of emotion and attention worldwide. IAPS provides experimental control in the selection of emotional stimuli, with its most significant strength being the ability to draw comparisons between results from different studies. IAPS is covered in more detail in Section 2.1.14.

The existing benefits of the IAPS are its collections of normatively rated affective stimuli:

- This allows for better experimental control in the selection of emotional stimuli.
- It facilitates the comparison of results across different studies conducted in the same or different laboratories.
- It encourages and allows exact replications within, and across, research labs that assess basic and applied problems in psychological science.

The approach of this research will be to:

- Try to replicate the strengths of the IAPS while using video as the medium.
- Establish that this type of method can effectively gauge video content in the same manner as IAPS for pictures.

4.2 Aims

This study aims to take pre-existing and recognisable IAPS methods, and apply them to video content. This allows video content, which is a heavily utilised media element, to be utilised in the realm of affective research. The study will be a cornerstone and foundation for affective video studies, allowing for experimental rigour when testing video content:

- Reproduce the IAPS methodology while using video as the medium.
- Add affective ratings to video content.
- Create an easy and recognisable method to provide affective ratings on video content for use in prediction.
- To test whether these ratings are comparable with the IAPS counterparts

4.3 Analysis Methods

4.3.1 Participants

There were 48 participants for this study recruited on the local university campus, 34 males and 14 females, with an average age of 25. They were selected from communication and media, computing, and psychology students. Participants were required to be over the age of 18, and to be unaffected by flashing images to participate in the study. Before conducting this

study, ethical approval was obtained from Manchester Metropolitan University, and the EthOS application reference number is 48042.

Max Age	49
Min Age	19
Average Age	25
Standard Deviation	7.39
Median	22

Table 4:1 – Statistical overview of the participants of the study

4.3.2 Experimental overview

The original IAPS contains 1,182 standardised, internationally accessible, colour photographs in a wide range of semantic categories (Lang, 1997), including content such as eroticism and human mutilation. Any such images that might cause ethical issues were not included in this study, and have been removed from the originals, as they are outside the remit of this study. To filter out some of the ethical issues that the original IAPS content creates, 40 of the IAPS images were selected randomly and checked for ethical issues, until 40 that were not of ethical concern were picked.

Some of the types of content that were shown were based on the IAPS descriptive categories:

- Man in pool
- Women
- Clowns
- Lizards
- Dogs
- Buildings

These pictures formed the image resource bank for the affective video system (AVS). To construct the video resource bank, these images were then matched with a video. Next, the videos were matched based on sharing a similar visual relationship with the image selected. Furthermore, the sound was removed from the video to remain as close to the IAPS methodology as possible. Finally, the videos were presented randomly to each user to normalise the results.

The experiments were done in groups of 20 to 40 participants, depending on the uptake at the time of running the experiment in a computing lab. Firstly, participants were shown a short clip of 20 seconds, and then they were asked to provide feedback based on the presented stimuli. This feedback was then captured onto the SAM scale for all 40 clips for each user. This process is shown in Figure 4:1.

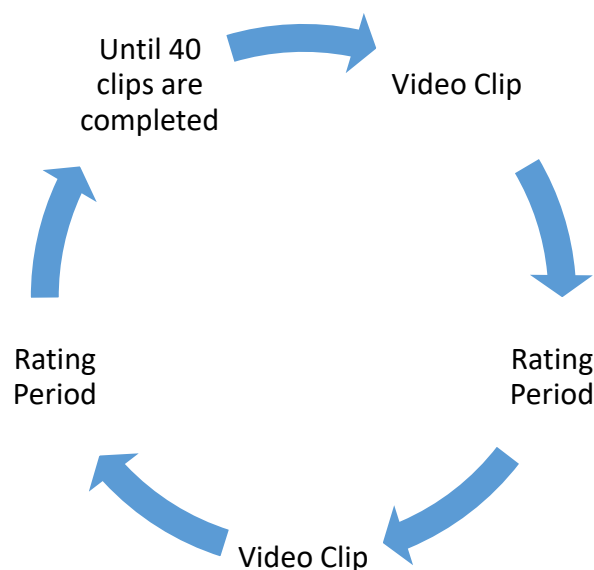


Figure 4:1 – AVS testing process

The experiment consisted of 40 clips from the AVS resource bank, and the user was asked to provide a rating for all 40. The experiment took 20–30 minutes, including a testing presentation before the trial run, to highlight critical areas for participants and to allow questions to be answered before the experiment. The trial run comprised three clips that took 2 to 3 minutes. This was an essential part of helping participants understand what was expected from them during the testing, and minimised participant errors in the study.

The videos were put onto YouTube and then embedded into a website populated via a database in random order. The users watched a clip and then provided their responses.

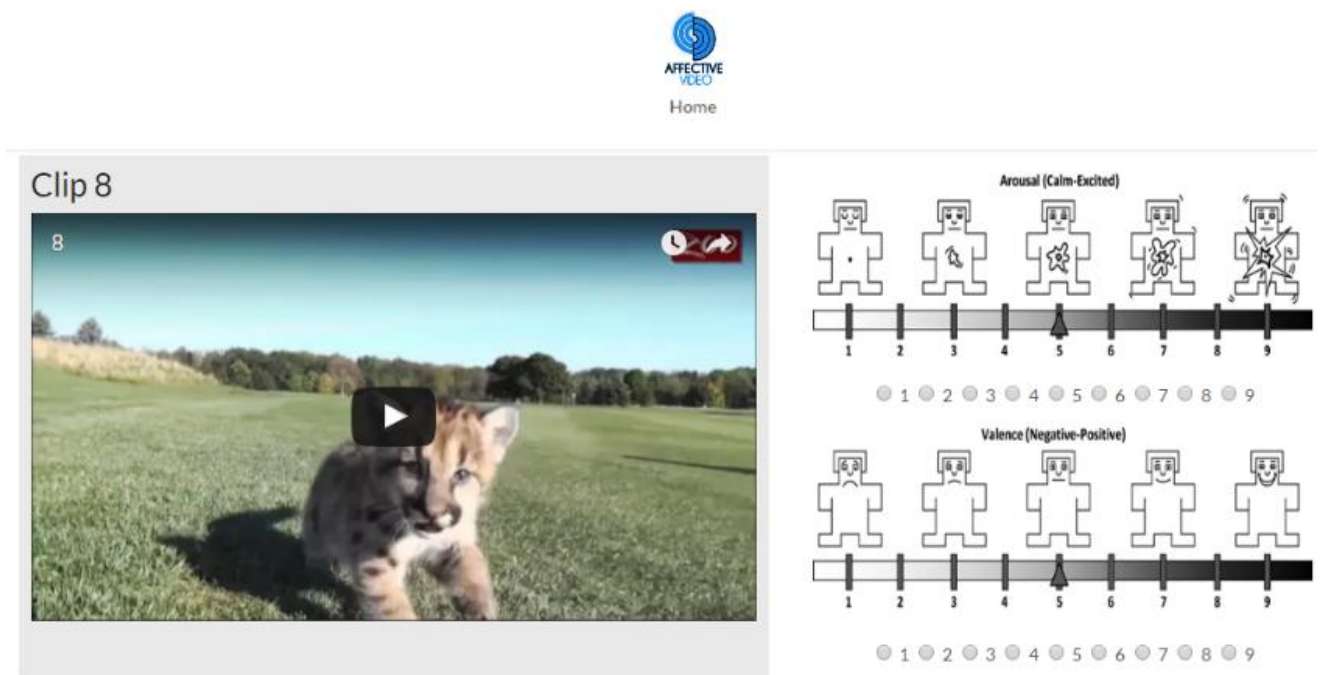


Figure 4:2 - AVS overview

4.3.3 Experimental procedure

This experiment adapted experimental procedures from the IAPS to the affective video system (AVS). The experiment substituted pictures from the IAPS for short videos showing

similar imagery to the original IAPS pictures. The experiment included an information sheet for participants, so that they were aware of the purpose of the study, as this would not influence the outcome. The information provided covered everything from the time taken to complete the study, to the full experimental procedures.

The steps provided for the user were:

1. The user watches clip—the user watches a clip that lasts 20 seconds
2. The User rates clip—the user provides a value for two scales provided below the video.
3. The user continues this process until all clips have been rated.

The test presentation part of the experiment was designed to help users fully understand what was expected of them during the experiment. This provided details on how to answer correctly using the scale of the emotion felt on each clip, and how strongly they experienced it (9 being very happy, 1 being very unhappy, and 5 being neutral). The clips used in testing were different and were only used for testing purposes.

The main experiment included all 40 clips from the AVS video content, where the user was required to provide a rating for all the clips. Each clip lasted 20 seconds; once the clip finished, the user provided the feedback required per the rating procedure.

The recommendations given to participants were that; accuracy was more important than speed; there were no right or wrong answers; they should provide an honest response; they should focus throughout the experiment, and finally, they should remain silent.

4.3.3.1 Video content

The video content included in the experiments was 1920 × 1080 resolution and was embedded onto an HTML webpage that comprised the complete set of video content for the experiment.

4.3.3.2 Rating procedure

The rating system for the experiment was the self-assessment manikin (SAM) (Bradley & Lang, 1994), which measures pleasure and arousal view section 2.1.13 for further information.

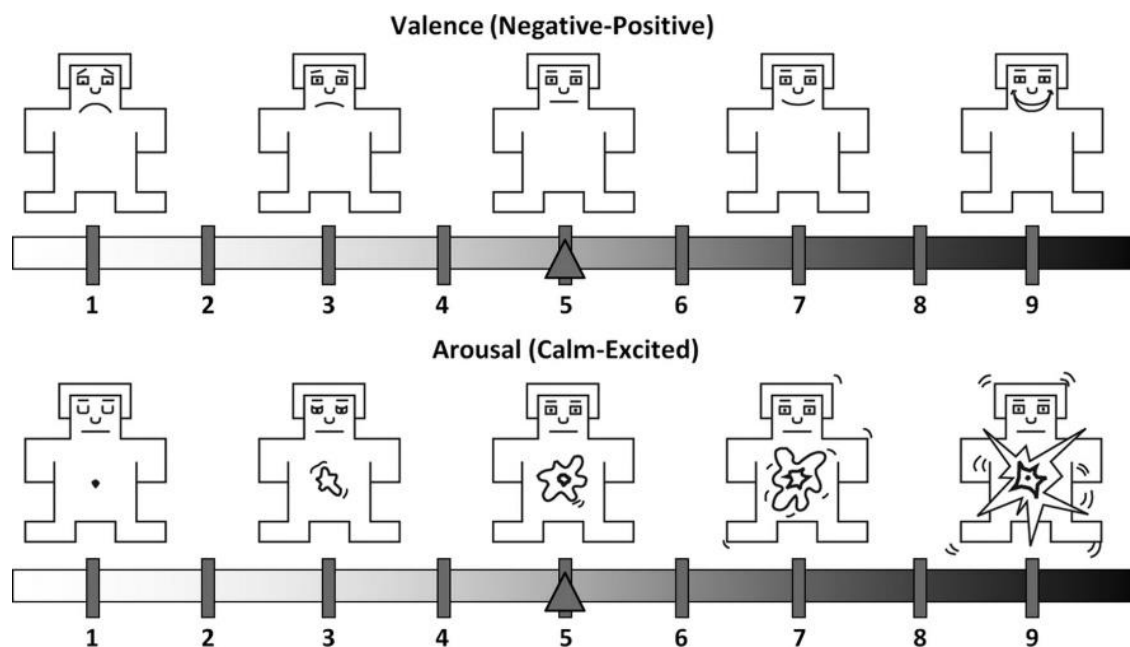


Figure 4:3 - The self-assessment manikin (Bradley & Lang, 1994).

This scale allows participants to indicate feeling happy or unhappy, or calm or excited, using the midpoint of each scale as neutral. Participants can provide an indication on the 1–9 point rating scale for each dimension. After viewing a clip, the study participants rated how they felt the emotional impact based on the SAM scale, providing a value for valence and arousal.

4.3.3.3 Content classification

The model shown in Figure 4:4 is one of the approaches used to interpret natural human expressions of combinations of underlying emotions. This model is designed to automate the processing of these emotions and reflect the processing of a user's emotions. One of the strengths of this method is its sensitivity to the selection of the base classifier. Base classifiers are probabilistic classifiers based on applying Bayes' theorem, with strong independence assumptions between the features. Their benefits are that they are extremely scalable (Mower, et al., 2011).

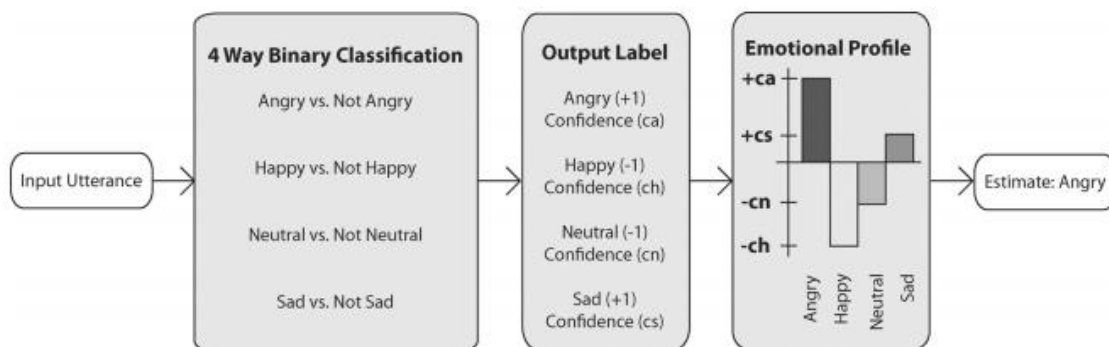


Figure 4:4 - Four-way binary classification (Mower et al., 2011)

The strength of this model is its straightforward approach, utilising a binary classification for each emotion to provide a best-fit outcome. This system's methodology showed great promise for a simple emotional recognition system, and was expected to work effectively with the affective methods outlined in this report.

With the above research in mind, the experimental results were plotted against the circle model, as shown in Figure 4:5, once the results were averaged. This showed where each clip

sits on the AVS, and a comparison could be drawn on whether there is a relationship between AVS and IAPS.

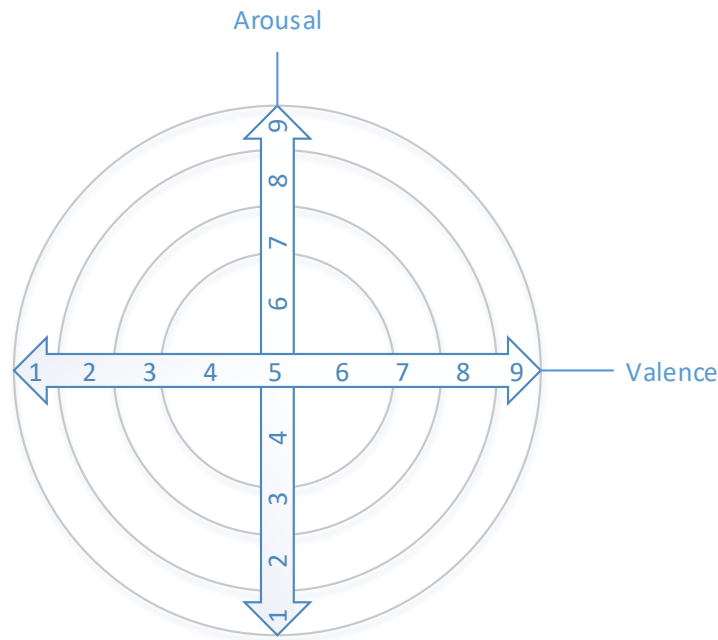


Figure 4:5 - Example of how SAM scores map to the circumplex model

4.4 Results

This section presents the study's findings, providing a direct statistical comparison against the IAPS and the new AVS dataset.

	IAPS Valence Values	IAPS Arousal Values	AVS Valence Values	AVS Arousal Values
Max	8.21	6.47	7.50	6.38
Min	3.04	2.79	2.90	2.52
Standard Deviation	1.42	1.00	1.00	0.87
Average	5.69	4.83	5.34	4.61
Median	5.87	4.69	5.60	4.63

Table 4:2 – Descriptive statistics for the IAPS vs AVS

The overall mean ratings for the selected IAPS data used within this study were 4.83 (SD – 1.01) for arousal and 5.69 (SD – 1.44) for valence. The following are the mean ratings for the AVS: 4.61 (SD – 0.88) for arousal and 5.34 (SD – 1.01) for valence.

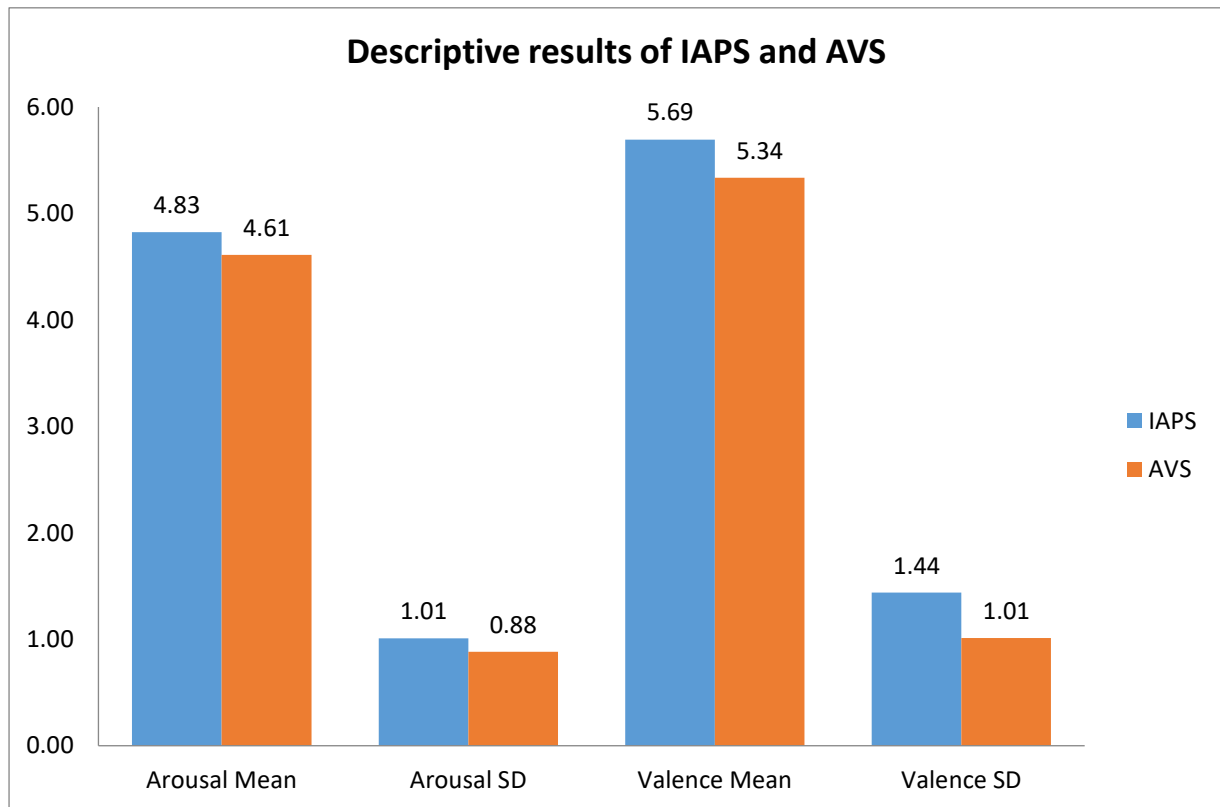


Figure 4:6 - Graphic representation of descriptive statistics for the IAPS vs AVS

Figure 4:6 shows that the descriptive statistics are very closely linked. There was only a 0.36 difference between the arousal means, and only a 0.21 difference between the valence means. Furthermore, the difference between the arousal standard deviation was only 0.13, and for the valence, it was a 0.43 difference. This provides a clear picture of the relationship between these two datasets.

4.4.1 AVS dataset

Appendix A shows the dataset that was constructed from our experiments and upon which the results are constructed

4.4.2 Paired t-test

The purpose of the paired t-test was to determine whether there is statistical evidence that the mean difference between paired observations is significantly different.

We ran a paired t-test with a significance level of $\alpha = 0.05$. The result for arousal was a p-value of 0.0953, showing that the difference is considered not to be quite as statistically significant. On the other hand, the paired t-test results between valence were a p-value of 0.0082, showing that the difference is considered very statistically significant.

4.4.3 Pearson correlation results

We then ran Pearson's correlation to measure how strong the relationship between the two variables was. The results were as follows: The Arousal results showed a value of R of 0.662. This is a moderate positive correlation, which means there is a tendency for high X variable scores to go with high Y variable scores (and vice versa). The value of R^2 , the coefficient of determination, is 0.438. The p-value of the correlation is < 0.00001 . The result is significant at $p < 0.05$. Figure 4:7 shows the results for the Pearson correlation graph, which indicates that there is a similarity between the two regression equations. They are both linear and have coefficients with relatively minor differences.

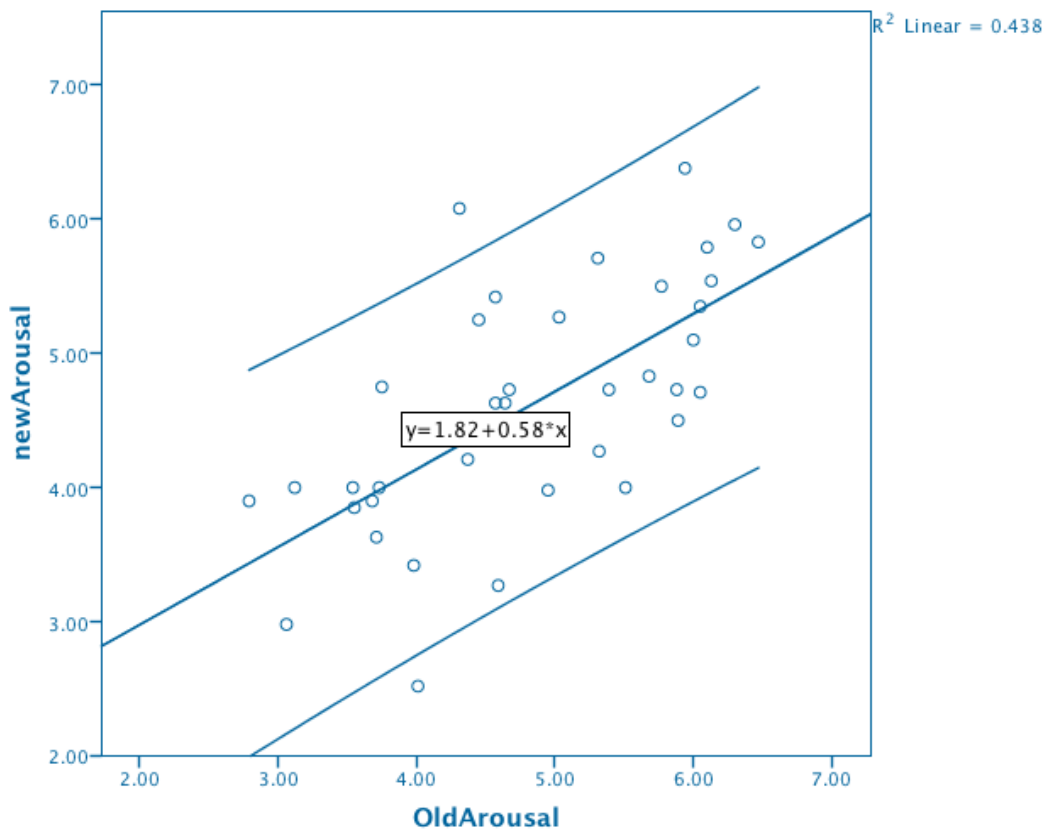


Figure 4:7 – Pearson Correlation Graph for arousal

The valence results showed a value of R is 0.8353. This is a strong positive correlation. The value of R^2 , the coefficient of determination, is 0.698. The p-value of the correlation is < 0.00001 . The result is significant at $\alpha < 0.05$. Figure 4:8 shows the results for the Pearson correlation graph, which indicates that there is a similarity between the two regression equations. They are both linear and have coefficients with relatively minor differences.

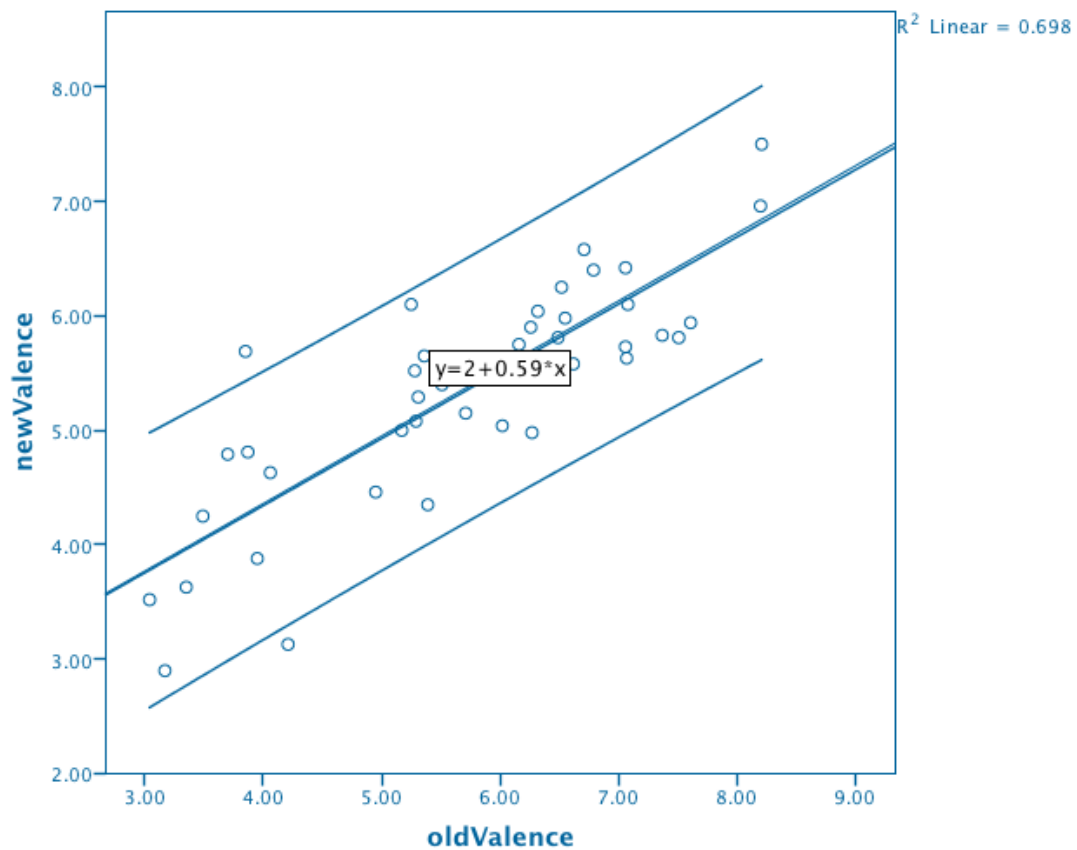


Figure 4:8 - Pearson Correlation Graph for valence

4.4.4 Affective space comparison

This section will focus on an affective space analysis based on the four categories presented in (Yazdani, et al., 2013), which is just a way to interpret the four key quadrants of the circumplex model. This provides insight into how this data maps within the affective space and provides an emotional inference. Figure 4:9 shows an overview of these categorisations to provide an initial insight between the two datasets, which indicates that there were 21 clips in the IAPS and 20 in the AVS for Pleasure, there were 8 clips in the IAPS and 8 in the AVS for Joy, there were 2 clips in the IAPS and 7 in the AVS for Sadness, and there were 9 in the IAPS and 5 in the AVS for Anger.

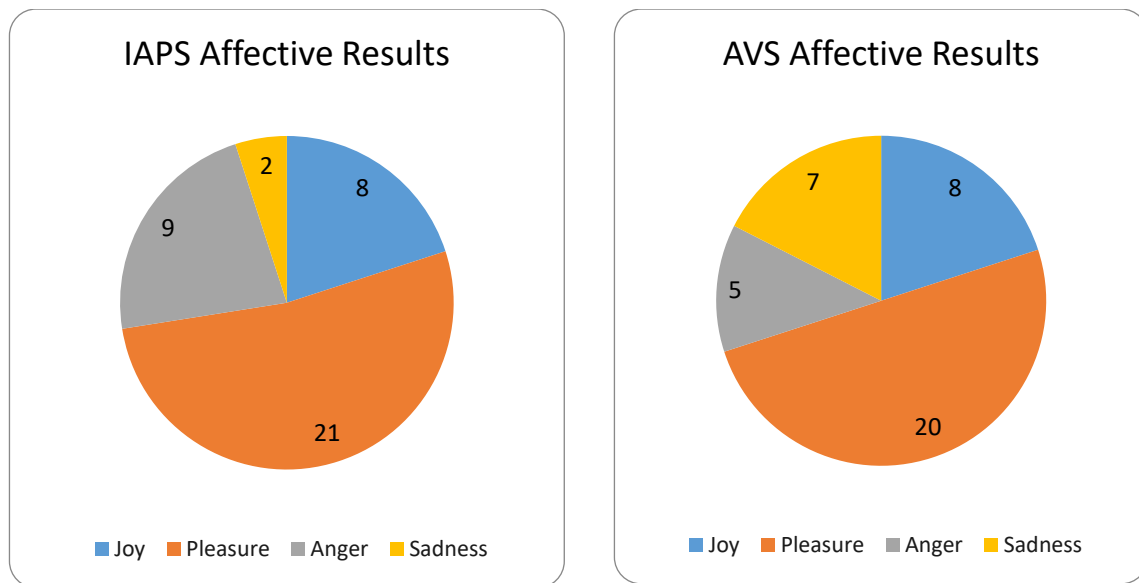
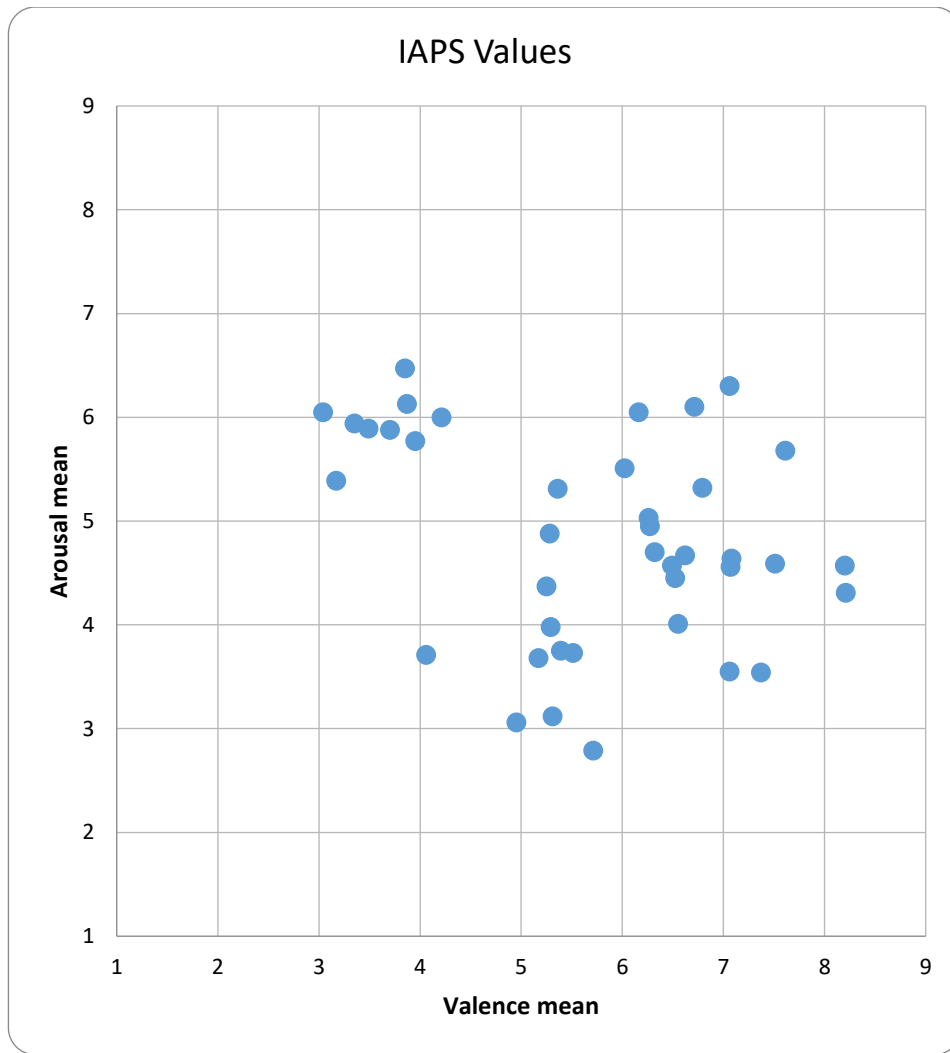


Figure 4:9 – Overview of affective space results between the IAPS and AVS

4.4.5 IAPS results

Figure 4:10 shows the original IAPS results mapped into an affective space, to provide insight into how the original dataset looked for comparison.



4.4.6 AVS results

Figure 4:11 shows the AVS results mapped into the affective space in relation to the average scoring provided by users for arousal and valence.

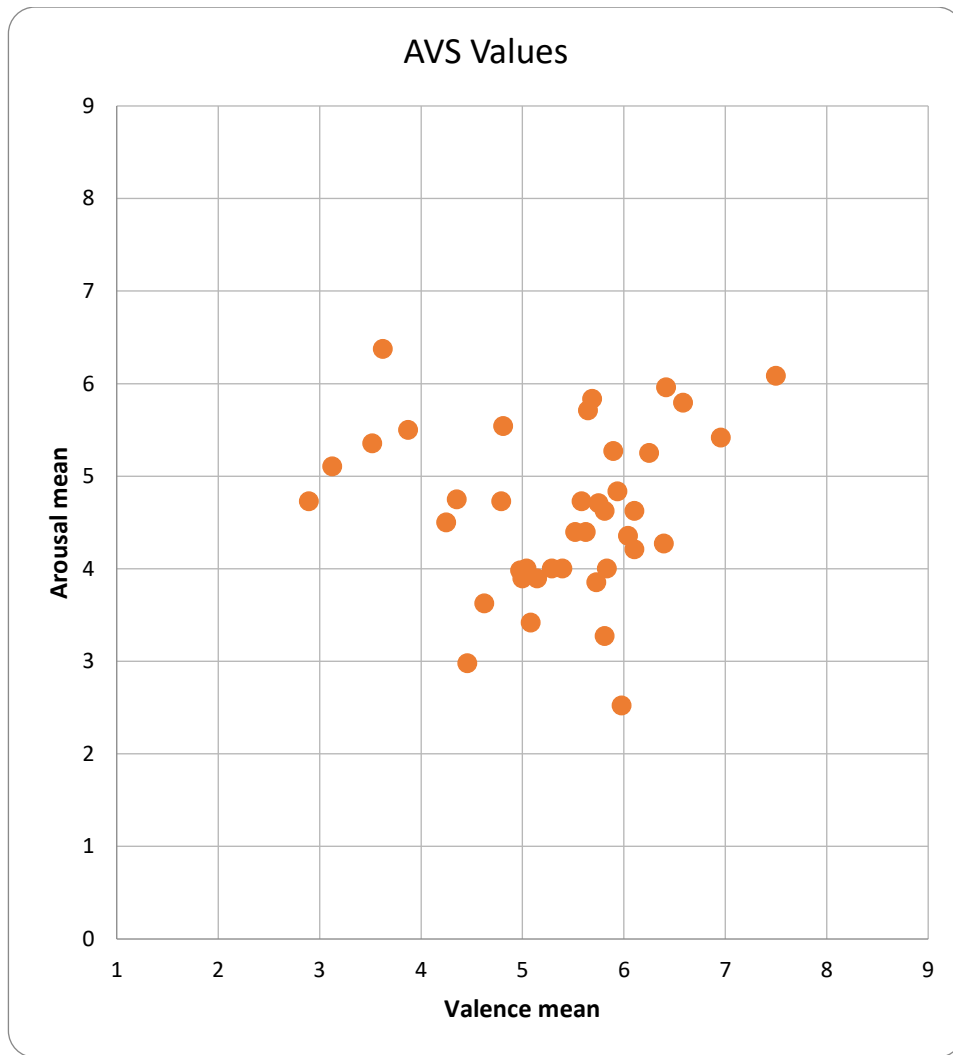


Figure 4:11 - AVS Results

From Figure 4:10 and Figure 4:11, we can see a similar grouping between the two. However, the AVS results were more tightly compact when compared with the IAPS. The study results have replicated a similar distribution to the original IAPS study when comparing the 40 original values to the 40 new ones. The affective space was determined by valence and arousal from the average score for the items across all participants.

4.5 Discussion

The above results have analysed whether we can transfer a method in IAPS onto video content and determined whether this method is still valid when you change the content type. The first discussion points will be broader since there was no one-to-one mapping, meaning the results were not exactly the same.

This study aimed to replicate well-defined pre-existing methods to carry into further experimentation; however, there is a difference between a statically presented image and its counterparts in video content, which is multiple images. So, there was never an expected outcome of a one-to-one mapping, even based on this assumption alone. It was, however, assumed that they would be sufficiently similar statistically to still allow them to be considered relevant to proceed with future studies.

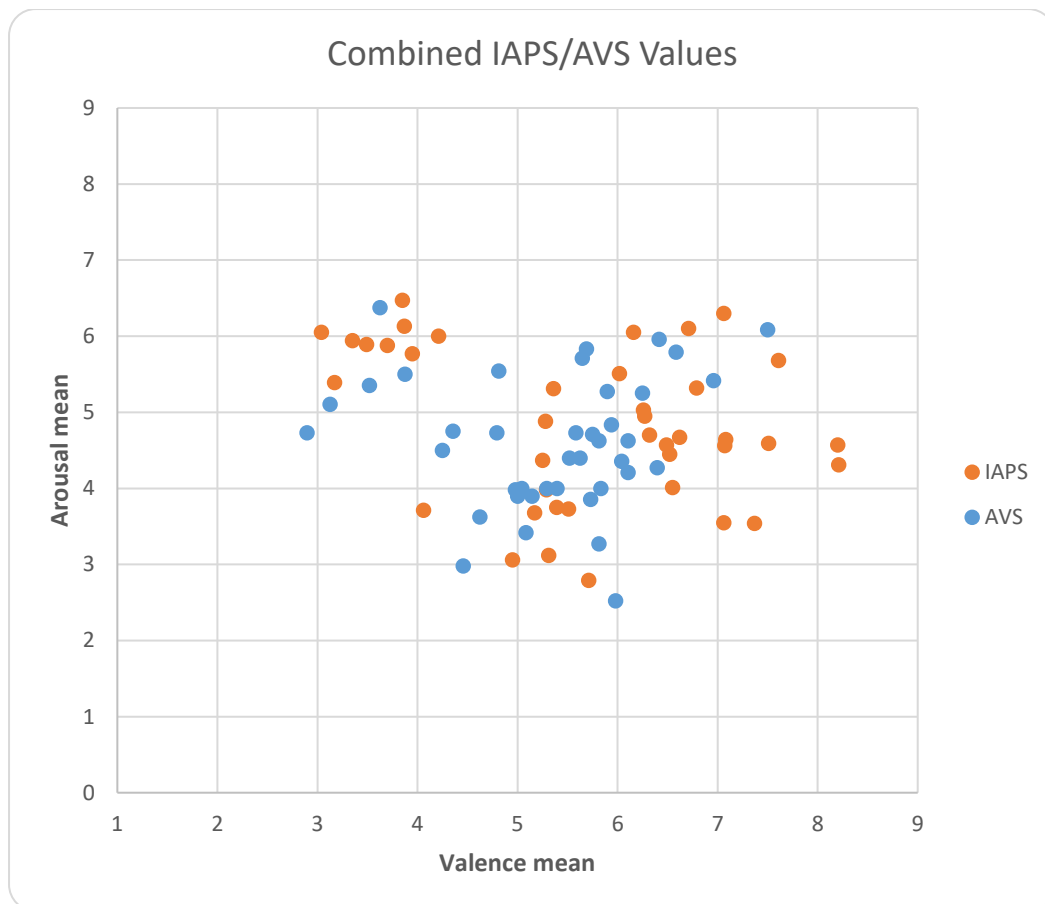


Figure 4:12 - Overlay Values for IAPS and AVS. The affective space has been gauged on arousal and valence dimensions.

Figure 4:12 shows that the values are clustered close together between the original results and the AVS results. The average difference between the points when compared with the IAPS results is 1.06, with 20 points less than 1 and 20 points more than 1. As the rating scale provided a range from 1 to 9, this means that 50% of the points could be considered within the same scoring metrics as the original data, and 50% of them would fall just outside.

Further discussion and research would be required to finalise the following discussion; however, the current format is presented for future reference as a significant trail of thought that can be followed.

4.6 Conclusion

The original aims of this study were to take the pre-existing and recognisable IAPS methods and apply them to video content, providing future works with a cornerstone and foundation for affective video studies, and allowing for experimental rigour when testing video content. Based on the results presented, this study met this aim and has validated that IAPS methods can be transferred to video content and still provide results that can be considered valuable, in the same manner as the pictures used in their studies.

Another objective of this study was to reproduce the IAPS methodology while changing the content from picture to video. This was undoubtedly achieved during the study, and was proven to still provide significant results that can be used as affective ratings for video content.

As the AVS methodology is essentially an extension of the IAPS methodology to allow IAPS methods to be used with video content, this was certainly achieved due to the well-founded IAPS methods being easily transferable to video content.

Finally, the most significant contribution of this study was to test rather than assume that the IAPS methods would successfully transfer from picture-based content to video, which has been proven by the results presented.

4.7 Summary

This chapter's focus was to use the well-known predefined methodology in IAPS and to test its effectiveness when transferring its methodology to video content, thus testing the assumption that it would successfully transfer, regardless of the content change. Then, section 4.3 shows the methods used to facilitate the change in purpose and repurpose the IAPS for video content. Finally, the statistical results provided in Section 4.4 prove that this is the case, which means that we can move into the next phase of this research.

Section 4.5 presented the discussion on topics around the research that led to Section 4.6, concluding that the method could be successfully transferred to video content and utilised in future studies.

4.8 Progression

By concluding that this method can be used, the next research question to be explored is which features will best allow us to map affect, and how these can be extracted from video content to be used in a system at a later date. This means the subsequent study will investigate what users deem to be the most important features using the repertory grid approach.

Chapter 5 - Perceptions of Video Features Using a Repertory Grid

5.1 Introduction

This research intends to aid in developing new technological mechanisms to catalogue, find, and access movies based on emotions. This could help to assess the videos' emotional impact and find films or scenes that induce the selected emotion in the viewer, and whether it was the intended emotion of the production. In addition, this study will help identify the important perceptual features to users that can be statistically quantified features for extraction for the regression analysis.

Repertory grids are used to understand the relationship between constructs and a series of elements of interest. Constructs are descriptive terms used to define a user's experience and are subjective to that person. The person chooses these descriptive words without prompting or leading consciously or unconsciously. These words highlight the implicit theory about the events from within the person making those decisions (Fransella, 2004)

This chapter examines emotional responses to video by gathering feedback from the participants, using the repertory grid methodology, after viewing a range of video content. Users were asked to describe distinguishing factors from the video clips; this investigatory technique can be utilised for video evaluation. The study will also provide an overview of the key visual and audio features perceived in the content of various film genres, leading to the formation of a generalised group view.

5.2 Aims

The research in this chapter aims to obtain insight into the key features in video clips that viewers perceive, which allow them to differentiate between them.

The research objectives are as follows:

- Use repertory grid interviews to gain insight into the audio and visual characteristics participants identify, which allows them to distinguish between films of different genres.
- Provide insight into the subjective experiences indicated by users of a video content system.
- To try to identify which features are most frequent and important to users.

5.3 Repertory grid

This research utilises the repertory grid, a tool based on Personal Construct Psychology (PCP), attributed to Kelly (1955).

This method is based on descriptions of the occurrences that are encountered by humans from experiences and by validating descriptions in their minds. Repertory grids can provide an overview of the subjective nature of a given user's experiences.

Personal construct theory (PCT) works on the principle that individuals build an arrangement of constructs that they use to make sense of the given task forming a worldview. The theory is that behaviour is governed by constructs, which have a bipolar dimension (Marsden & Littler, 2000).

PCT is the concept of bipolarity and provides insight into how users view the world. The theory states that people construe elements in the world as 'similar to' some things and 'different from' other things. This method does not require thinking or feeling but forces selective behaviours to make a choice. The interesting aspect of this method is that it takes place at many levels of awareness, including unawareness (Fransella, 2004), which is why this approach is well suited to affective research.

An overview of the main components of the repertory grid approach is as follows.

Domain is the topic of interest that needs to be defined. For this study, the domain was the perceptual features of film clips.

Elements are the things that are being examined with a particular method. In this study, these were the film clips.

Constructs are the dimensions of what is being examined. Here, they had to be bipolar; each element was related to one pole or the other of the given construct.

The technique of induction utilised within the repertory grid method is to ask participants to compare three elements chosen at random and then to comments on their similarities or differences

Linkages are the ways in which each element is described in terms of each construct. This study used a card sorting technique (Wood & Wood, 2008) with the film names on a scale of 1 to 5 on a whiteboard. The user placed the film to their score; this was then recorded. This allowed users to visually map their results for ease (Rogers & Ryals, 2007).

One of the other issues relates directly to the richness of video content, which is constructed of numerous images as well as audio stimuli. These two elements, working in conjunction with each other, produce media-rich content.

5.4 Analysis Methods

A repertory grid was used as a recognised research method, and it is a structured approach to gathering personal constructs from individuals. It allowed the participants to build up their own repertory grid during the interview used to elicit constructs.

The research procedure was as follows:

1. The participant watched all eight video clips.
2. The participant was then asked questions facilitated by a repertory grid interview technique during a one-to-one interview to devise constructs and rate the clips according to those constructs.
3. The participant and interviewer repeated this process until they could not provide any more feedback.

Before conducting this study, ethical approval was obtained from Manchester Metropolitan University, and the EthOS application reference number is 48044.

5.4.1.1 Repertory grid method overview

This section provides a more detailed account of the experimental procedures followed while collecting this data from the participants.

5.4.1.2 Example Repertory grid as drawn on a whiteboard

The first tool utilised for this experiment was a grid that was drawn on a whiteboard; it helped the participants to understand and grasp the concepts behind the repertory grid approach. It presented the 1 to 5 scale for the scoring and the ability down the right-hand and left-hand sides to add a given construct against which to rate film clips.

	1	2	3	4	5	
<i>Constructs</i>						<i>Constructs</i>

Figure 5:1 - Example repertory grid from experiment

Another tool utilised was eight cards shown in Figure 5:2 that represented each of the eight film clips being rated. These were used in combination with the whiteboard shown in Figure 5:1 to allow an easy mechanism for the users to present their perceptions.

The eight film trailers were chosen randomly from the IMDb charts for top-rated movies (IMDb, 2021). This list of 250 movies is rated the best on the IMDb website. A ninth clip from the original eight had to be randomly selected, again due to the trailer for ‘Closer’ potentially raising ethical issues.

The video content was selected to represent something that participants recognised from everyday life, as opposed to the bespoke content used in Chapter 4, where the experiment content is directly related to the IAPS images. Furthermore, since the problem area of this research is concerned with using affective methods to filter commercial film content, this led to the conclusion that film trailers would be a good source.

The language was kept explicitly to English, as the study was predominantly tested in the UK and used the English language for communication with participants. Figure 6:1 provides an overview of the genres covered by the AVS dataset, based on the current genre categories offered on the IMDb website.

Anchorman 2	Jurassic World
The Dark Knight Rises	12 Years a Slave
Transformers—Age of Extinction	Johnny English Reborn
Titanic	World War Z

Figure 5:2 - Card's example (cut up individually)

5.4.1.3 Repertory grid experimental procedure

The experiment follows the repertory grid approach. Participants reacted to the elements (film clips) and constructs (explained in bipolar terms of like and dislike).

The next phase of the experiment was for the user to watch all the film clips. The second phase was to explain the method in more detail with “like” and “dislike”, which constitutes the test phase to get the user accustomed to the rating process and the use of the whiteboard and cards.

By this stage of the experiment, participants were familiar with the procedure and the tools of the methodology. A technique known as “triads” was utilised, where three cards (elements) are selected randomly, and the participants choose the two they perceive to be the most similar. This method questions the reason behind the grouping and the differences between them to provide a bipolar construct. Once the bipolar construct has been elicited, the participant is handed all the cards (with the names of the film clips) to provide a rating 1 to 5 scale based on the construct provided previously.

Steps used during the experiments:

1. Pick three cards at random
2. Question to ask: “Out of the three elements chosen, which two seem to have something more in common with each other?”
 - 2.1. Why did you group those together?

2.2. What are the characteristics of the two you have grouped together?

2.3. What word would you use to describe the opposite of the word selected (giving the bipolar for the word selected in 2.2)

3. Allow the user to write the construct (no longer than one or two words)
4. Place the original three cards on the whiteboard
5. Give them the rest of the cards to rate on the 1–5 construct scale
6. Once the rating phase is finished, ask if they are happy with the results

The results are then noted for the given construct for all eight films, and then the above steps are repeated until the participant can no longer provide constructs. Once all participant's data had been gathered, they were joined together to form a single grid. Once this process is complete, it provides a matrix of ratings, which is the repertory grid for the participant and can be used to analyse the results.

5.4.2 Participants

Fifteen participants were recruited on the Glyndwr University campus to undertake this study. This number was chosen since other repertory grid studies have recruited similar numbers for their research. For example, a study conducted by (Matthijs Kwak, 2014) operated with 18 repertory interviews, and another study conducted by (Marc Hassenzahl, 2000) used 11 repertory interviews. Whilst a more extensive sample was possible, researchers of such works have noted that limiting recruitment to 15 to 20 participants

would have been adequate to generate sufficient constructs to build the required datasets (Felix B. Tan, 2002).

Participants were a mixture of students and lecturers selected from the university's media, computing, and psychology disciplines. In addition, 113 constructs were generated from this experiment for the eight-film clips.

Potential participants were provided with an information sheet before the experiment to provide information about the study before their participation, as well as being able to ask any questions or any points of clarification that were needed.

The participants had to meet the following criteria to be eligible to participate:

- Be over the age of 18.
- Have no medical condition preventing them from participating and not be sensitive to strobe lights/images.

5.4.3 Experimental methodology

The users watched eight clips, each of which was between two and three minutes. Once the participants had seen the content, they were asked questions about it and asked to provide a 1 to 5 rating of the content for a given construct. This method was based on the Repertory Grid covered in 5.3.

The experiment was conducted in a home-like environment to replicate a real environment the participants may experience in their day-to-day lives.

5.4.4 Web interface

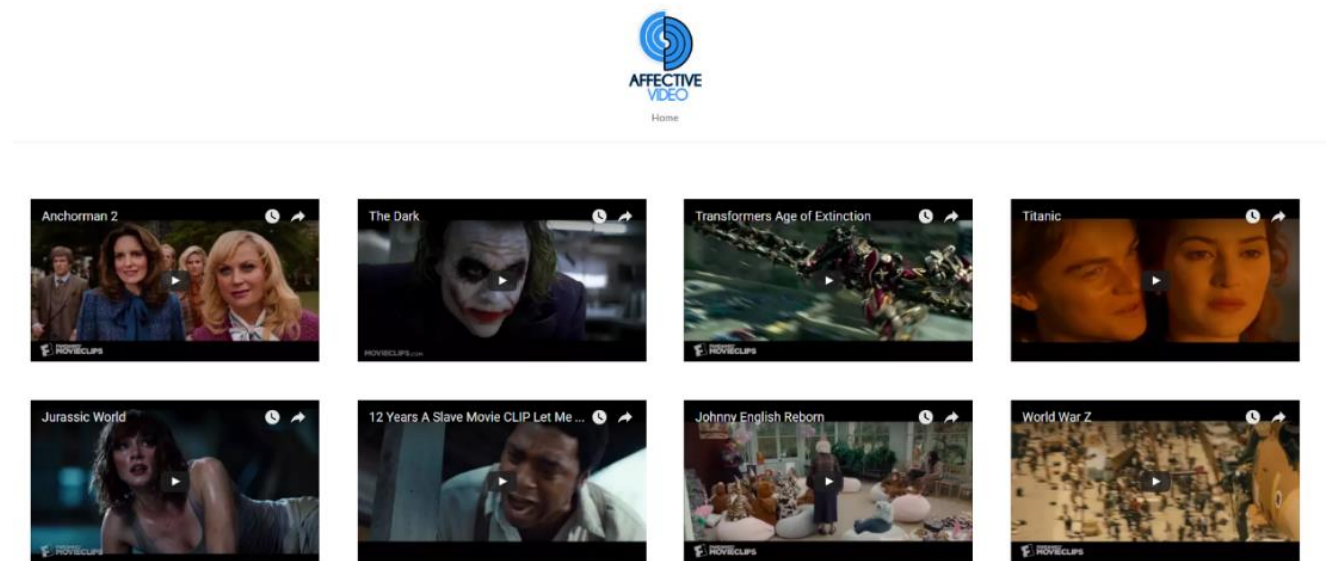


Figure 5:3 – The system used to present video content to the user.

Figure 5:3 shows the web interface used to show the user the video content. The user would work through the content and put each clip into a full-screen view as they progressed.

The video content for this study was clips from the following films in Table 5:1, including the associated genres from the IMDb website (IMDb, 2021). The number of film clips was based on how many would typically be viewed at the cinema before a featured film.

Films were randomly selected from the IMDb website, and the clips were selected based on them being recognisable to the participants, which takes the study into a more real-world application compared with the content used in Chapter 4 experiments directly relating to the IAPS images. There were also additional benefits of the content being freely accessible and as close to films without users having to sit for several hours to provide feedback on multiple films. Film clips suited the nature of this study as they provided much quicker affective scoring.

Film	IMDb Genre 1	IMDb Genre 2	IMDb Genre 3
Anchorman 2	Comedy		
The Dark Knight Rises	Action	Crime	Drama
Transformers--Age of Extinction	Action	Adventure	Sci-Fi
Titanic	Drama	Romance	
Jurassic World	Action	Adventure	Sci-Fi
12 Years a Slave	Biography	Drama	History
Johnny English Reborn	Action	Adventure	Comedy
World War Z	Action	Adventure	Horror

Table 5:1 - Film clips included in the Repertory Grid study with IMDb Genres (IMDb, 2021)

5.5 Results

This study generated a large amount of qualitative and narrative data relating to the participants' explanations of elicited constructs. In addition, a vast amount of qualitative data was analysed to identify emerging themes. This data was analysed using the Web Grid Plus (Plus, 2021), and the results were as follows.

5.5.1 Principal Component Grid

Figure 5:4 shows a Principal Component Grid, which treats the elements as points on a Cartesian coordinate system where the dimensions are the first two components of the PCA on the horizontal and vertical axes, and then each of the constructs and elements as they are scored on those two dimensions.

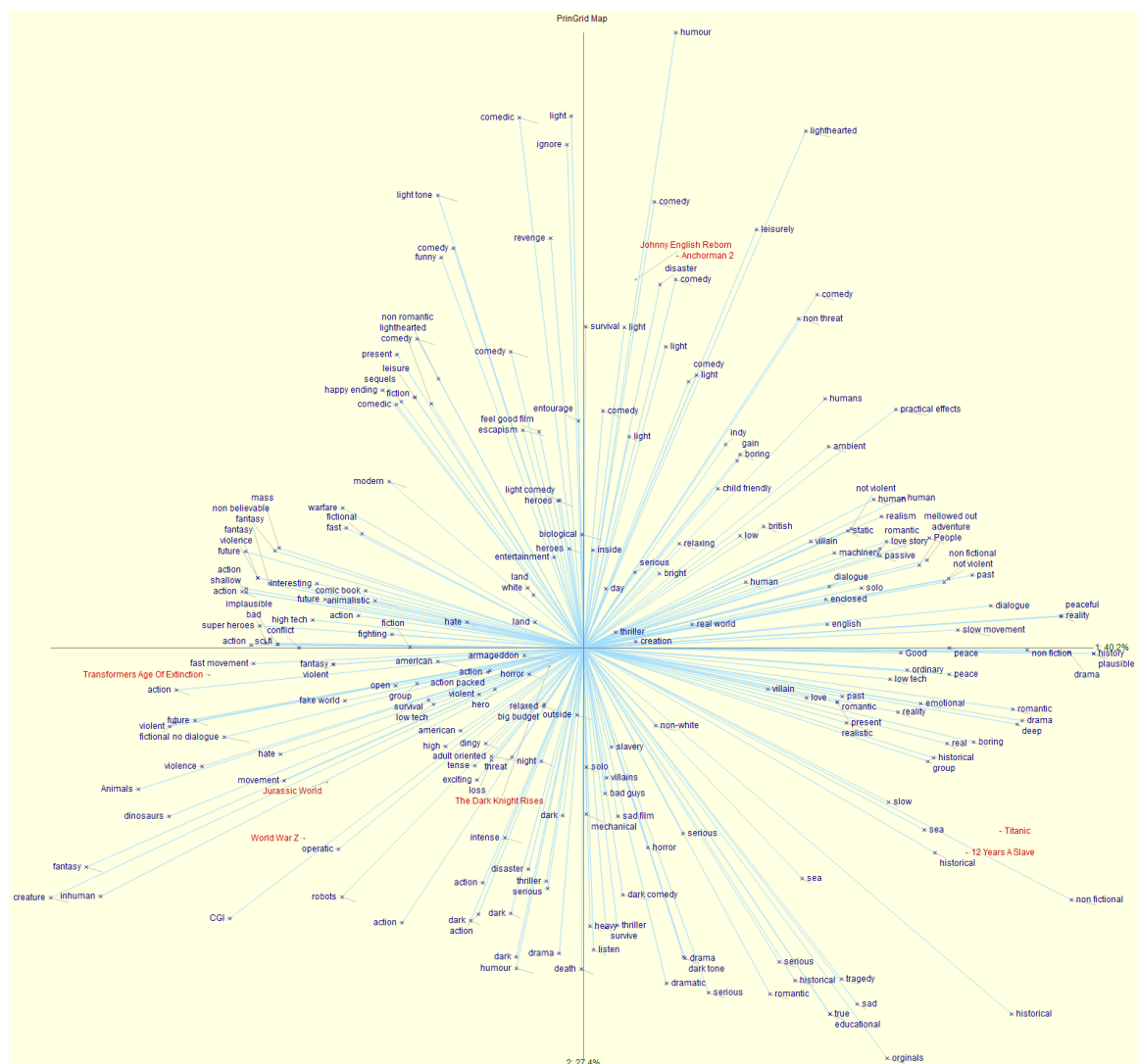


Figure 5:4 - Principal Component Grid Map

Figure 5:4 provides insight into what constructs map to what elements. There are numerous examples of suitable matches based on the positioning of the film clip and the constructs used in this study. An example of this is the element *Jurassic World*, which is within a relatable distance to the constructs ‘dinosaur’, ‘fantasy’, ‘fake world’, creature, and ‘inhuman’. Two other examples are *Johnny English Reborn* and *Anchorman Two*, both of which had numerous constructs relating to ‘comedy’, ‘light-hearted’, ‘humour’, and ‘light’. The repertory grid method seems to be able to map descriptive constructs to elements that would be considered easily relatable.

One important distinction is that a repetition of some words, such as ‘light’, can be defined in several ways. This is one of the observations during the study as participation grew. For example, some people referenced ‘light’ in terms of ‘light’ and ‘dark’, which could be a reference to the setting of the film in relation to brightness, or ‘light’ and ‘dark’ in terms of the overarching theme, like ‘hero’ or ‘villain’. These types of subjective constructs provided an interesting dimension to this research because they highlighted where participants used the same word, but it may have had a completely different meaning.

5.5.2 Film clip results

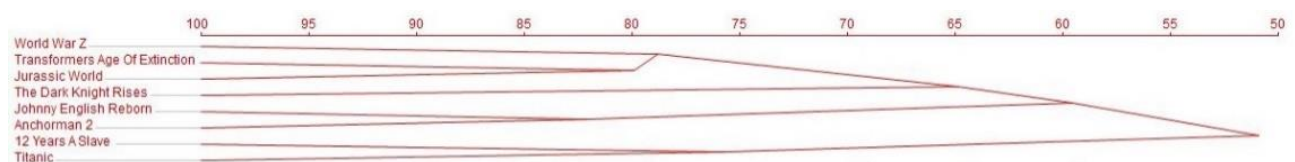


Figure 5:5 – Percentage of similarity between the content

Figure 5:5 shows a dendrogram that illustrates the hierarchical relationship between film clips. The shorter the line height, the more closely related the given objects were. We can

see that the two most closely related clips, based on the qualitative data from the users, were *Johnny English Reborn* and *Anchorman Two*. Both these clips can be considered under the genre ‘comedy’ when referencing Table 5:1. There is also a relationship between *World War Z*, *Transformers—Age of Extinction* and *Jurassic World*, which all share ‘action’ and ‘adventure’ as genre categories. Finally, *12 Years a Slave* and *Titanic*, which share the ‘drama’ genre, also seemed to be related. This provides an interesting insight into how users provided constructs that link relatable genres between the film clips. This would suggest that there is a causality between genre and perceivable content characteristics to the users.

5.5.3 Construct Analysis

Figure 5:6 shows a dendrogram for the relationships between constructs. This highlights the hierarchical relationships between constructs using hierarchical clustering, represented as a tree. The data is analysed in more detail in Table 5:2, Table 5:3 and Table 5:4.

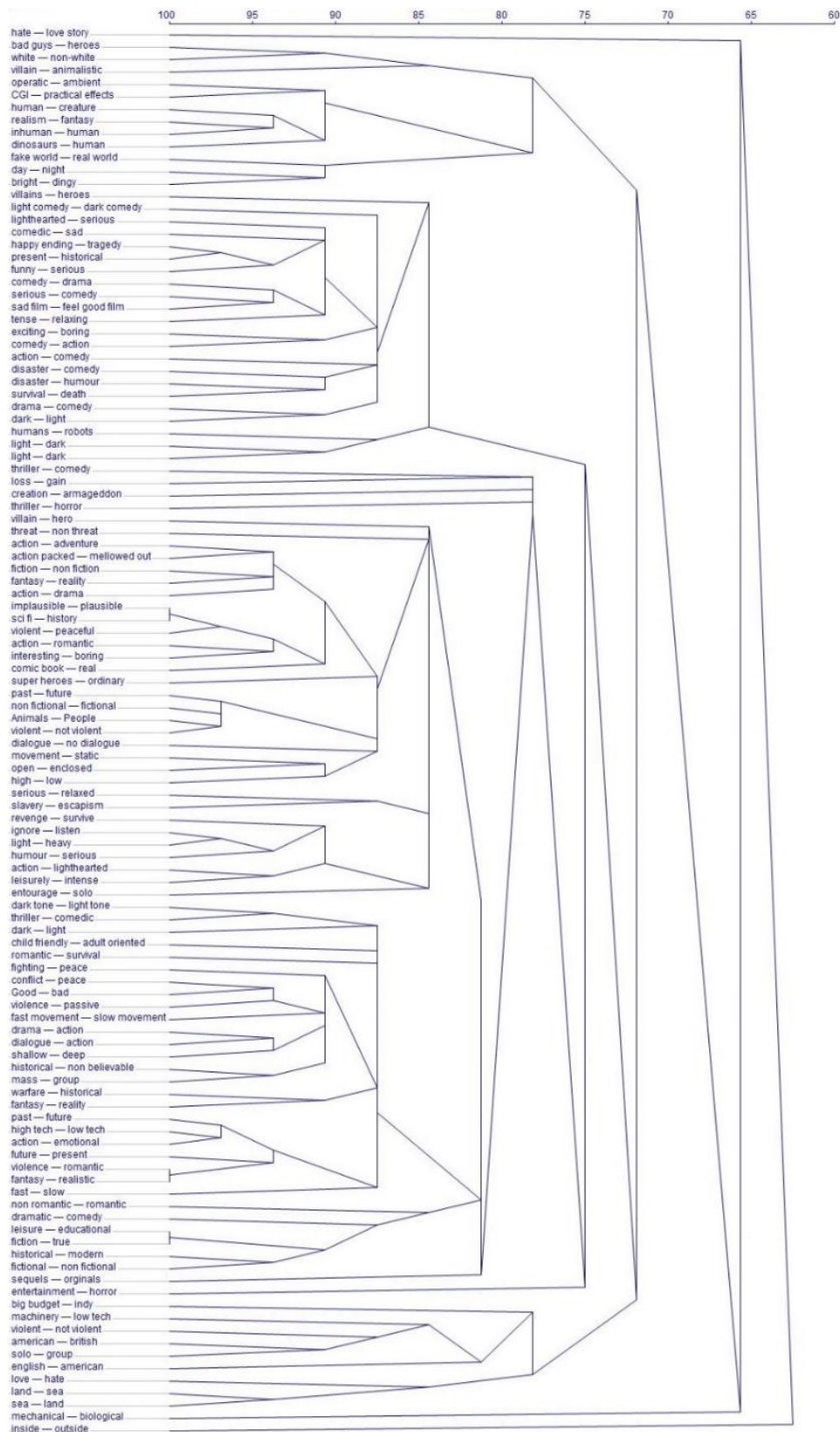


Figure 5:6 – Focus Cluster Dendrogram for Personal Constructs

5.5.4 Construct Analysis

The focus cluster method is shown in Figure 5:6 highlights constructs that were closely matched. Table 5:2, Table 5:3, and Table 5:4 show the most closely matched constructs provided by the participants. An example of two closely related constructs is ‘sci-fi—history’ and ‘implausible—plausible’, which, when you examine the semantic meaning, could be considered to be describing the same concept.

Constructs			M%
sci fi	—	history	100
implausible	—	plausible	
violence	—	romantic	
fantasy	—	realistic	
fiction	—	true	
leisure	—	educational	

Table 5:2 - 100% match for constructs

Constructs			M%
action	—	emotional	96.9
high tech	—	low tech	
non fictional	—	fictional	
past	—	future	
light	—	heavy	

ignore	—	listen	
present	—	historical	
happy ending	—	tragedy	
Animals	—	People	
violent	—	not violent	
violent	—	peaceful	
implausible	—	plausible	
violent	—	peaceful	
sci fi	—	history	

Table 5:3 - 96.9% match for constructs

Constructs			M%
Animals	—	People	96.9
non fictional	—	fictional	
high tech	—	low tech	
past	—	future	

Table 5:4 - 96.9% match for constructs

Table 5:2, Table 5:3, and Table 5:4 provide numerous insights into the closely related constructs resulting from the study. Whilst a number of these highly related constructs will not be typically considered as audio or video features, there is potential that they could be scored true or false. This could lead to them being used for implementation in future

affective systems. For example, ‘fictional’ and ‘true’, by definition, lead to a binary conclusion because they are either fictional or real.

5.5.5 Word analysis

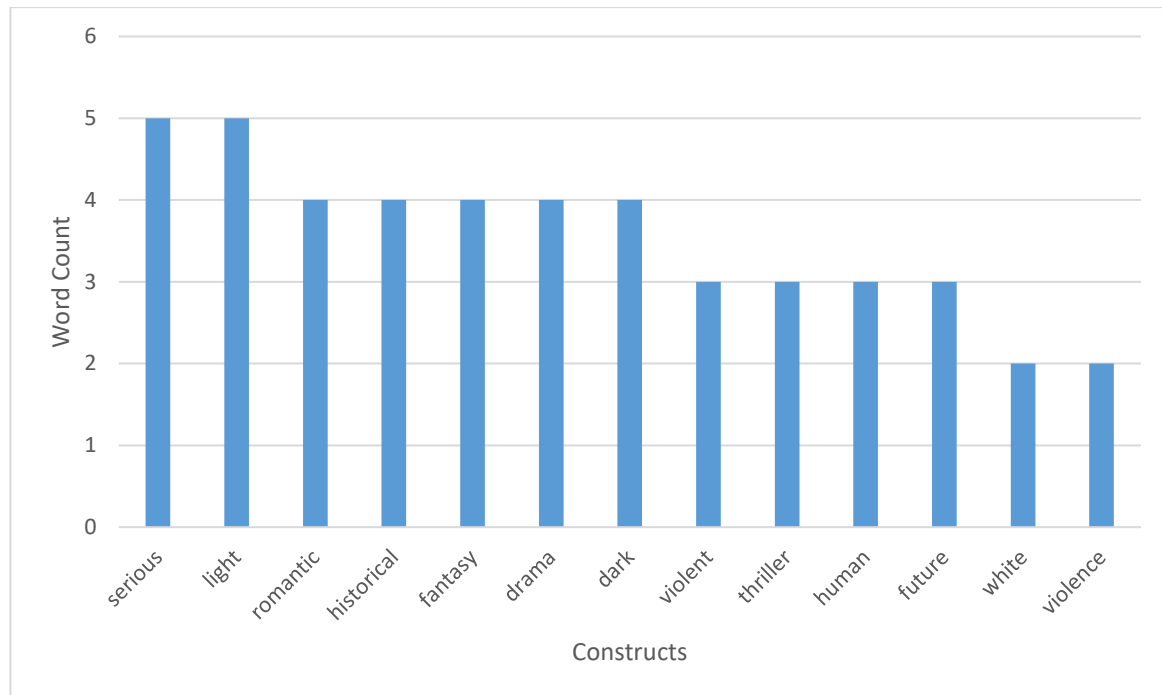


Figure 5:7 - Most recurring constructs

Figure 5:7 shows the most predominant word counts from the study, which can be observed as showing the recurring words provided by users for constructs. We can see numerous bipolar constructs such as ‘serious’ and ‘romantic’; ‘historical’ and ‘fantasy’; ‘dark’ and ‘light’. This figure highlights recurring themes for the users between separate interviews, providing insights into predominant user constructs.

5.6 Discussion

This chapter has provided exciting insights into users’ perceived features. First, we can see in Figure 5:5 transparent inter-genre relationships that make sense when we consider the

genre-based approach employed to date. Furthermore, this study has mapped semantic data to the genres provided by the participants in Figure 5:6, which highlights these relationships based on the constructs provided by users.

The results shown in Table 5:2, Table 5:3, and Table 5:4 highlight the strongest matching constructs, providing deep insights into perceived features that the end-user relates to and focuses on when determining between different film content. Unfortunately, a number of these constructs are closely related to the genre or content of the film clip based on the findings of this study.

One of the main criticisms of the repertory grid approach, is the low number of participants.

In section 5.4.2, several studies have been highlighted, showing that 15 participants are a robust sample. The repertory grid technique does not engage many participants simultaneously, as it is time-consuming. However, one of the greatest strengths of the repertory grid is that it combines qualitative and quantitative data. This means it is an excellent fit for this type of study, even with few valid and reliable participants.

5.7 Conclusion

The primary technique in this chapter was to use a repertory grid to provide insight into what features the users found important, and then to find out from this list of features which ones could be statistically benchmarked to be used with regression to provide the best outcome. Regarding the constructs and relation to features that can be used with regression, the outcome was 'light' and 'dark', which could be measured statistically, and 'white', which can be measured with RGB.

Other notable features were ‘action’, ‘comedy’, ‘historical’, ‘romantic’, ‘fantasy’, ‘drama’, and ‘thriller’, which all reference genres specifically based on semantic meaning. Another recurring theme was ‘real’ and ‘fictional’. This was represented with a few different constructs, showing that users distinguished between real and not real. Finally, there are several constructs that could potentially be extracted, such as ‘human’, ‘sea’ and ‘land’, which could be identified using feature extraction methods.

The original aims of this study were to gain insights into the key features presented in video clips that viewers perceive, which allowed them to differentiate between those clips. These insights were provided, and while it was not an exhaustive list of features that could be statistically extracted from the video content, it provided a certain direction on perceivable features to users using the repertory grid approach.

5.8 Summary

In summary, this chapter started with a clear aim of following on from the previous chapter by trying to build a foundational theory for features that are important to users, which may, in turn, lead us towards an affective video system. However, this study has also provided critical insight into users’ perception of video content, even though this differs from the original hypothesis. Finally, it has identified two potential features to carry into the subsequent study and valuable insights that could go beyond the scope of this research in terms of feature extraction.

5.9 Progression

This study has identified a list of audio and visual features to carry into the subsequent research, providing a sense of direction. As presented in the results, there were numerous references to 'light/dark'. While there is a question of semantic nature, this can be explored in the subsequent study by extracting this feature and running a regression analysis.

Chapter 6 - AVS Dataset Creation, AVS Data Analysis

6.1 Introduction

This chapter will examine the role of affective responses to video clips presented via our online system, which allows for gathering feedback directly from the users after viewing video content. Once the feedback has been collected, we can use the results to train predictive models to predict emotion/affective scoring in video content.

This study is based on the IAPS (International Affective Picture System) (Lang, 1997), widely used in experimental studies of emotion and attention worldwide. Its methodology provides experimental control in selecting emotional stimuli, facilitating the comparison of results across various studies.

This research expands on the IAPS, which comprises only images, whereas this AVS study uses video content, which is many images as video frames and audio; both video and audio in the scope of affective computing have many research aspects to comprehend covered in the section 6.2. In addition, these methods would allow for a new dimension for video prediction/categorisation.

Furthermore, other research studies have highlighted how the combination of arousal and valence can provide compelling results (Lang & Dietz, 1999). Another study investigated the influences of low-level video characteristics on users (Canini, 2012). Their study used low-level features relating to light, colour and saturation and created a method to automatically position a movie scene in the connotative space.

One of the main reasons for not using a pre-existing dataset was that there were limitations within existing datasets that would not meet all the requirements for this study. One of the main issues was a lack of “real world” content that would be accessible to everyone, and widely recognisable content. Therefore, a combination of the IMDb top 250 with their corresponding film trailers was selected, and it was essential to build a testbed dataset around previous works in Chapter 4. It was decided that film trailers were a much more suitable form of video content to display to users, as they are widely accessible and commonly viewed in most regions and by most cultures.

The new dataset is the most real-world affective video dataset of its type and quality. While LIRIS-ACCEDE and EMDb both used films for their video content, it was not mainstream video content. This study seeks to bridge that content gap and bring it in line with something the users may view in their everyday lives, leading to the distinction of a more “real world” application for the video content in this study. Additionally, this study was conducted online and outside of the lab environment. Participants were also using their own equipment in their own environment each time the results were gathered.

6.2 Aims

The work in this chapter aims to provide emotional ratings for video trailer content used in an affective machine-learning system. The study employed a method for affective video annotation based on the approach used in the IAPS, which was developed to provide a set of normative emotional stimuli for experimental investigations of emotion and attention for pictures.

The main aims of the study are as follows:

- Create an easy way to provide affective ratings on video content for use in prediction
- Add emotional ratings to video content to allow affective modelling
- Create an original dataset of emotionally rated video content.

6.3 Methods

The following section gives an overview of the methods used to collect affective data from users using the online AVS system. The study utilised online methods so that we could gather the required data during the COVID-19 pandemic.

The website system used for data collection was affectivevideo.co.uk utilising a web server built from PHP, HTML, JavaScript, and SQL with a back-end dashboard of WordPress. The participants were assigned a unique ID that they used to participate anonymously in the study.

6.3.1 Materials

There have been earlier investigations into the emotion-inciting properties of video content. For example, a study conducted by Tan (1995) investigated the universal affective responses of users in classic Hollywood films, as films are constructed to induce real emotions as the story unfolds. The study highlighted that users are genuinely engaged with the content, and the responses are a genuine reflection of affective characterisation from a given scene. This study was developed to provide a set of normative emotional stimuli for experimental investigations of emotion and attention.

Emotion is a fundamental building block when creating films. This is why filmmakers exploit this to induce the required emotion for their scene. It is also worth mentioning that film trailers, such as those used in this study, are also constructed with a specific purpose, typically to entice people to watch the film, not necessarily to induce specific emotions in the users. However, films classically fall into genres that loosely translate into the type of emotion they may wish to convey to the audience.

The film trailers were selected from imdb.com (IMDb, 2020) and then matched with an appropriate trailer on YouTube.com (YouTube, 2020). One hundred film trailers were selected from IMDb's top films covering a multitude of different genres. The language was specifically kept to English, as the study was predominantly tested in the US, UK, and other countries. Figure 6:1 provides an overview of the range of the genres covered by the AVS dataset, based on the current genre categories provided on the IMDb website.

The video clips were drawn for samples that could be viewed in everyday life, as opposed to the video content used in Chapter 4 experiments directly relating to the IAPS images. This led to the conclusion that film trailers would be a good source. These were selected from the IMDb charts for top-rated movies (IMDb, 2021), which is a list of 250 movies that have been rated the best on the IMDb website, and they were sorted by release date to provide the most up-to-date content, so that it was easier to get higher-quality video content.

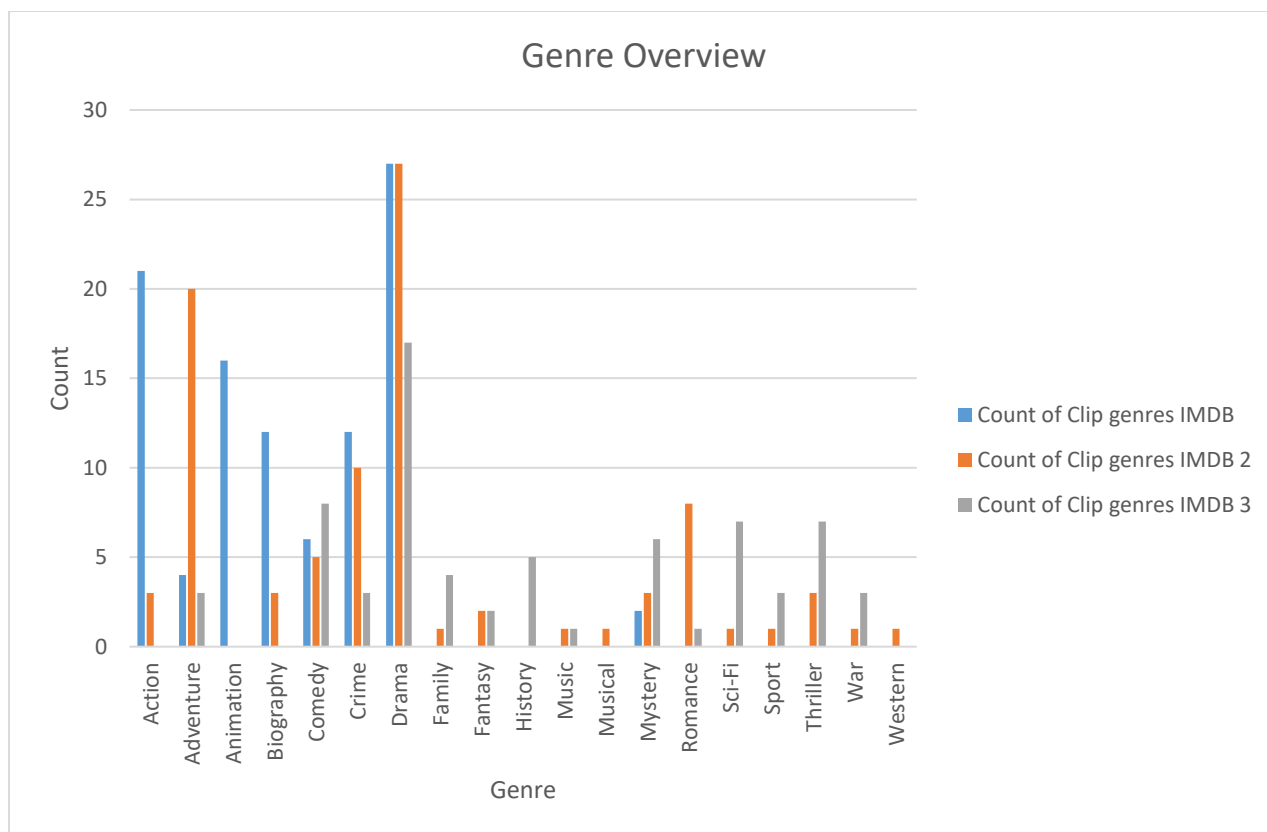


Figure 6:1 - AVS dataset genre composition

The study required users to be over the age of 18. This was an important ethical consideration for the project and was easier to enforce alongside rules adopted by YouTube. In addition, this provided a level of reassurance that participants would not be exposed to stimuli likely to cause an abnormal psychological disturbance.

One crucial ethical consideration was how to keep user data anonymous when using these methods. This was done by providing a unique ID that was not tied to any personal data for the user. Thus, the results had no personal data attached to them, meaning data could be published in its entirety where required.

Participation was primarily driven by Amazon Mechanical Turk (AMT), allowing the required number of participants to be engaged during the COVID-19 pandemic. Other studies that have utilised AMT are (Baveye, et al., 2015), (Chen, 2015), (Soleymani, et al., 2013) .

Each film clip has over 40 users' affective annotations between arousal and valence.

Participants were paid \$2.50 for annotating 10 clips, and they had to have spent over 20 minutes doing the study to be approved. This helped to filter out users who did not follow the experimental procedure. It was also a requirement that users who participated in the study were AMT "masters," who are workers who have proven "excellent" in previous AMT activities. Participants also needed to have a 95% approval rate for their tasks. Both of these requirements led to more consistent data collection and of a higher quality. We also limited the location to the US, UK and Canada. This was because the selected trailers were predominantly English, and the content was broadly popular in Western cultures.

The other aspect of participation was via social media, which accounted for users from countries not included above. This was the initial starting point of the study, but it did not gather the required traction to collect the required data. This meant we had to tweak the study and re-apply for ethical approval to use AMT. These tweaks included removing email from the capture on the website in line with Amazon's terms and changing the outreach method to AMT rather than social media.

6.3.2 Participants

The participants who entered the study were from 15 different countries. Most of the workers were from the USA (74%), with 18% deriving from the UK and 8% from other

countries; an overview is provided in Figure 6:4. Figure 6:2 and Figure 6:3 provide an overview of the scope of this study geographically, which shows that the study could be conducted on a more international scale in the future. Before conducting this study, ethical approval was obtained from Manchester Metropolitan University, and the EthOS application reference number is 14456.

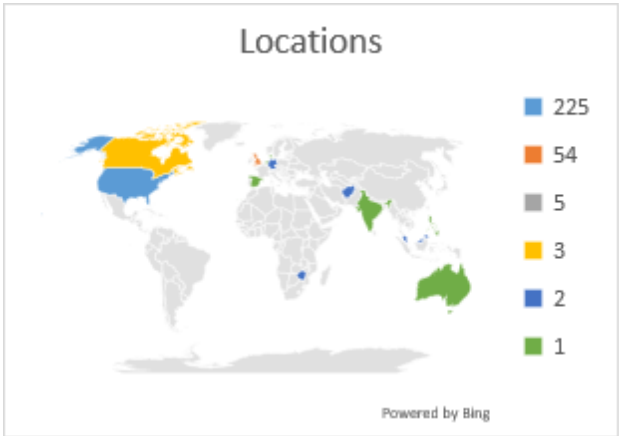


Figure 6:2 - AVS Data Collection Locations Overview

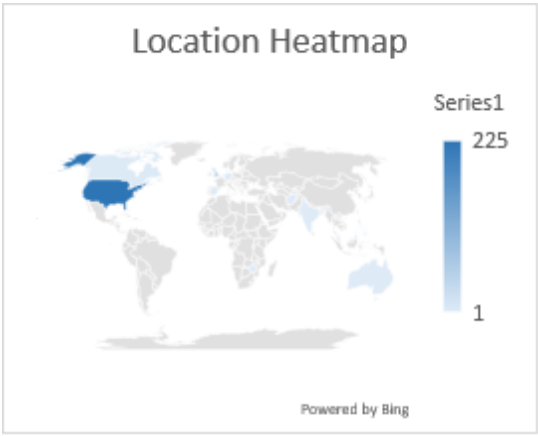


Figure 6:3 - AVS Data Collection Locations Heatmap

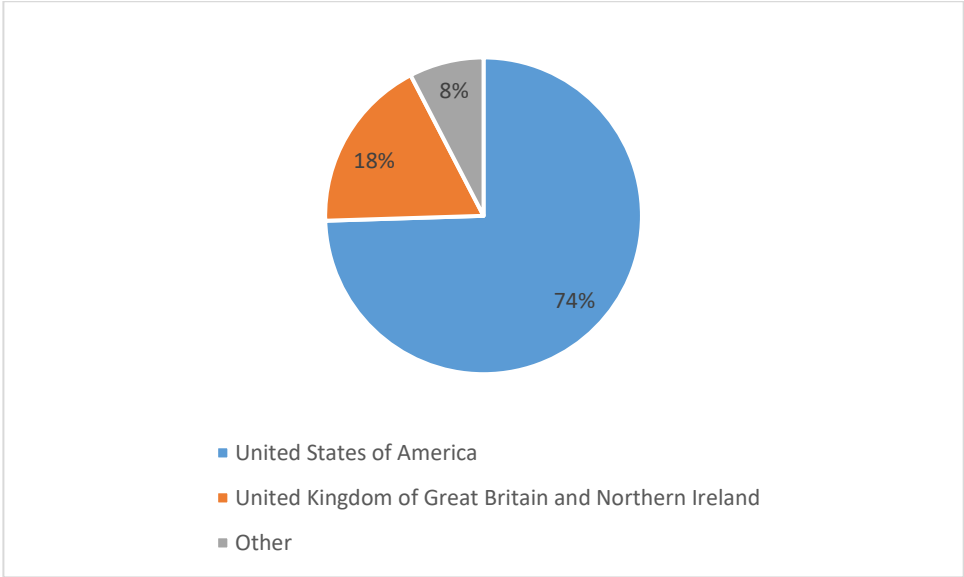


Figure 6:4 - Overview of Participation by country

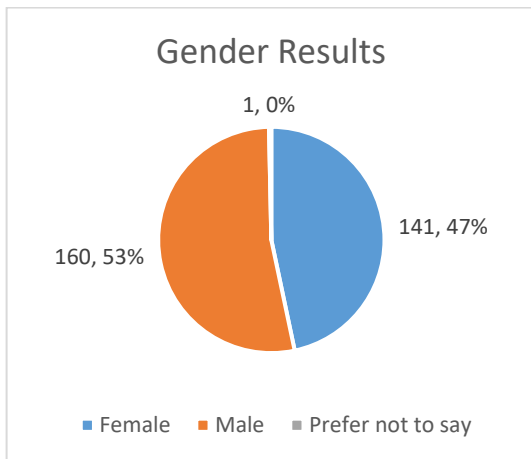


Figure 6:5 - Gender Overview

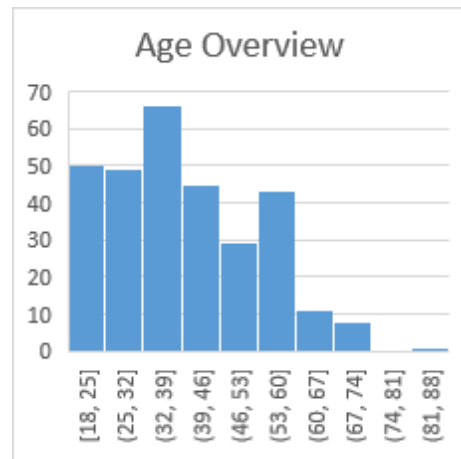


Figure 6:6 - Age Overview

160 male and 141 female participants, with one gender abstention, registered to participate in the study. This highlights the study's 53% male uptake, and a 47% female uptake study was representational of gender highlighted in Figure 6:5. Figure 6:6 shows a histogram age overview for the study, stressing that, once again, there is a diverse range.

Each user had to confirm that they were over the age of 18 to be eligible to participate in the study. The only other requirement for participation was to confirm that they had normal hearing and sight (which included sight corrected with glasses) for their age. Below are the questions that required answers before participation:

- Their age had to be over 18 to participate
- They confirmed that they had read and understood the information sheet

- They confirmed that they had understood the relevant sections of the data collected from them during the study, and that the researchers could look at it
- They considered themselves to have normal hearing and sight for their age
- They had to agree to participate in the study of their own free will without duress or pressure.

The study gathered 6202 unique arousal and valence ratings from participants, who each provided ratings for 10 video clips.

6.3.3 Rating mechanism (the SAM scale)

The film clips were rated via our online system, shown in Figure 6:9, using the SAM scale to provide the affective dataset in the 2D valence-arousal space. This two-dimensional rating approach has been preferred to other ways of classifying emotions into discrete clusters. However, such methods do not explore the complexity or diversity of emotions that could be induced over the time span of video content (Russell, 1980). After viewing a clip, users had to rate how they felt its emotional impact, using the SAM scale (Bradley & Lang, 1994), which is covered in more detail in section 2.1.13. This meant providing a value for valence and arousal. Next, the users watched the video clip, which was located above the SAM scale. Once they had watched the clip, they could provide their scoring on the SAM scale.

The online rating system developed was used to capture the results to help prevent any errors that paper-based methods presented. In addition, some systems were introduced to help with consistency within the study. For example, validation was added, for which the users had to enter results in all required fields before submitting their results.

The users entered the value based on the 1–9 scale to rate the value of the induced emotion, meaning how the video made them feel. The SAM scales' non-verbal design meant it was the perfect tool for this experiment, as it is considered easy to use regardless of age, ethnicity, culture, and education (Baveye, et al., 2015).

6.3.4 Experiment design

The experiment's design was built around the need of gathering crowdsourced data. This was seen as the most obvious choice to achieve this study's goal and has been used in other similar studies to build any new datasets covered in Section 6.3.1. However, this requires a set number of responses to be considered viable. In this study, we collected 6202 ratings that had to be individually collected from users using the SAM scale.

Another important aspect of the experimental design was presenting simple and digestible steps to the users to avoid confusing them, but to be informative and not hinder their progress.

The following steps were highlighted to the participants:

Step 1: Potential participants will read over the information page sheet

Step 2: Participants can then register by providing age, gender, location, email, and consent

Step 3: Participants are provided with a link to the experimental test page

Step 4: Participants are passed through to the experimental page on completion of the testing phase

Step 5: Participants watch a clip that lasts 1-4 minutes

Step 6: Participants rate the clip on the SAM scale

Step 7: Participants continue this process until all 10 clips have been rated

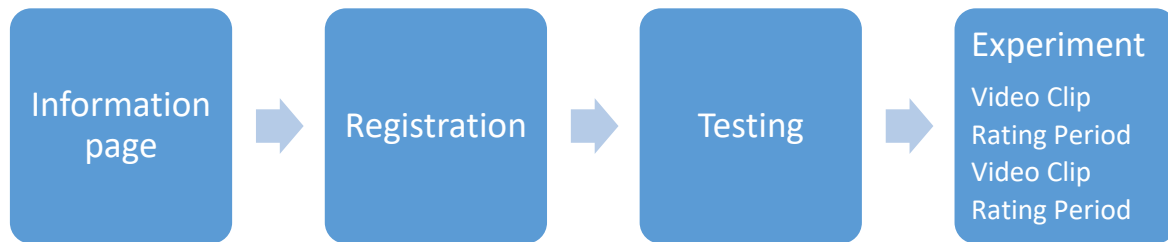


Figure 6:7 – Experimental Procedure Overview

6.3.4.1 Testing overview

Practice Rating

This is a practice to help you get a feel for how the ratings are done. Your ratings for this video clip will not be stored.

Some of the video clips you watch may prompt emotional experiences; others may seem relatively neutral. Your rating of each video clip should reflect your immediate personal experience, and no more. Please rate each one **AS YOU ACTUALLY FELT WHILE YOU WATCHED THE VIDEO CLIP**.

What do you need to do?

- Step 1. Watch the clip
- Step 2. Provide rating on the SAM scales

Clip 58



Arousal (Calm-Excited)

1	2	3	4	5	6	7	8	9
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Valence (Negative-Positive)

--	--	--	--	--	--	--	--	--

Figure 6:8 - AVS Testing Overview

To ensure consistency, this method and several protocols were implemented. The first protocol was an in-depth information sheet that provided all necessary information to the participant before conducting the experiment. A testing stage was introduced in which three video clips were required

to be scored using the SAM scale before entering the live experiment. This ensured that the participants were familiar with what was required of them before providing real data.

6.3.4.2 Experiment overview

It was decided to break the 100 clips into 10 batches of randomly allocated clips. This meant there was no selection bias, and each user was presented with 10 unique clips to score. Ten AMTs were run separately, and no user could rate more than four batches during the course of data collection.

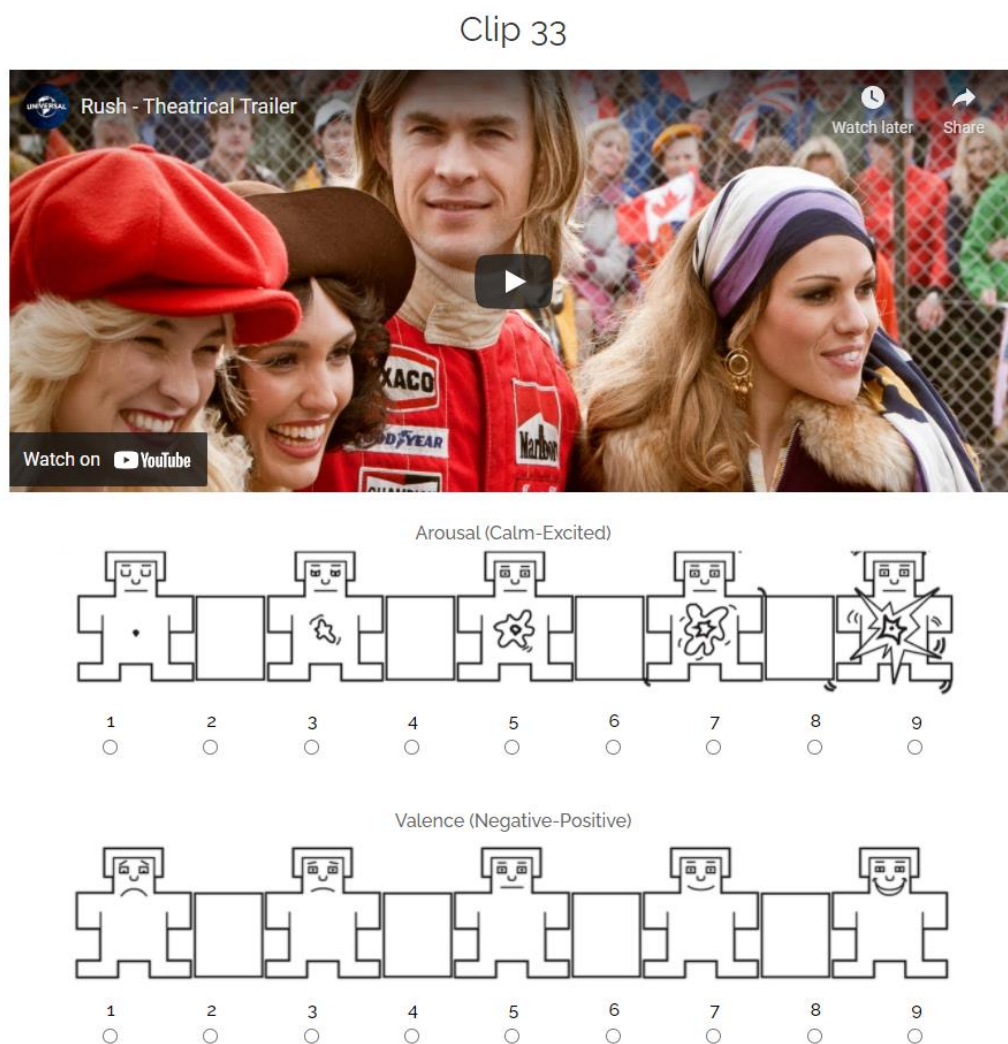


Figure 6:9 - Example of how users were presented with the clips during the AVS study

6.4 Results

Figure 6:10 shows that the ratings of the video clips largely fell within the ‘joy-affective’ space, with 70% of the clips falling in this affective categorisation. Again, this is an exciting insight into the affective scoring that film trailers may convey to users.

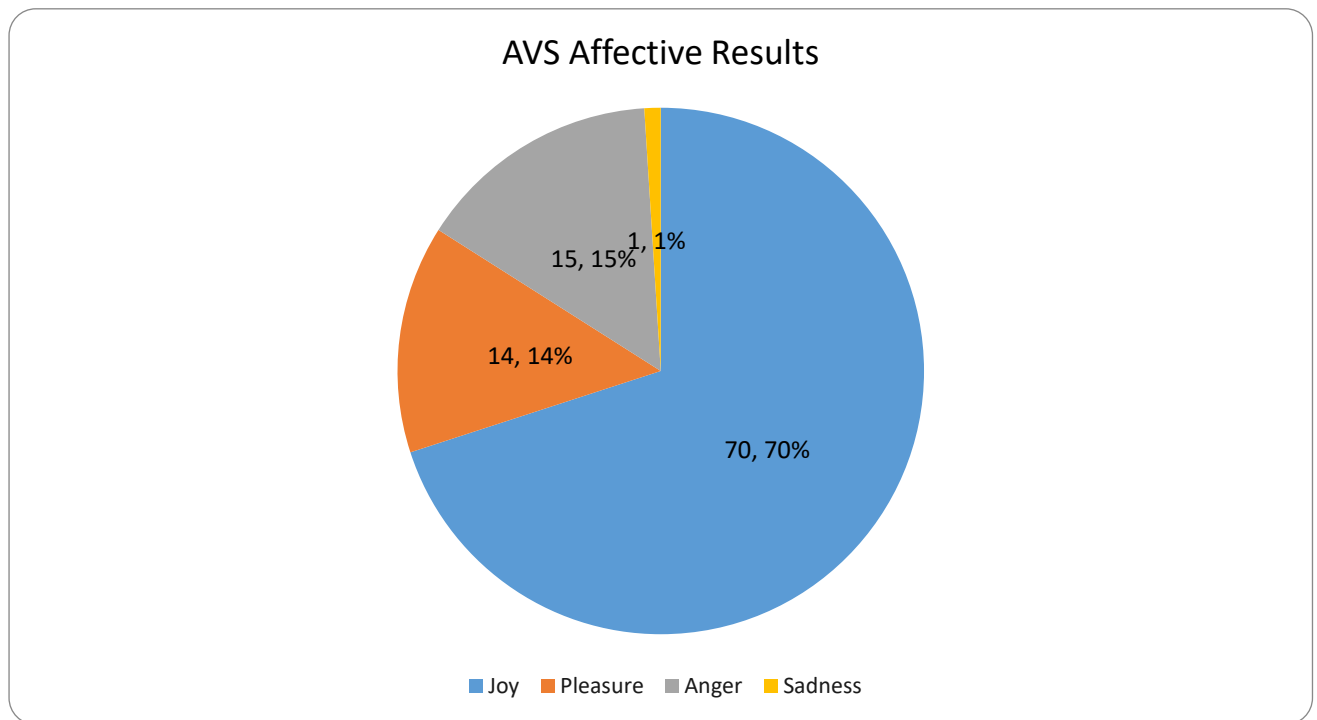


Figure 6:10 - An affective overview of the video content

6.4.1 AVS descriptive statistics

Table 6:1 shows the descriptive statistics for the AVS dataset.

	Average Arousal Values	Average Valence Values	Standard Deviation Arousal Values	Standard Deviation Valence Values
Max	7.23	7.48	2.46	2.68
Min	3.26	3.91	1.42	1.21
Standard Deviation	0.80	0.75	0.25	0.29
Average	5.79	5.71	1.94	1.90
Median	5.87	5.72	1.93	1.86

Table 6:1 - Descriptive statistics for the AVS data

6.4.2 AVS film clips results

Appendix B shows the average and standard deviation results for the AVS dataset in the same style as the IAPS results.

6.4.3 Distribution of ratings

Figure 6:11 shows the study's average arousal and valence of the user data. Again, we can see from the presented graph that there is a clear distinction towards the top right-hand side of the affective space.

As an observation, the data points are quite closely clustered. It is also interesting to note that there was a similar pattern in a study conducted by (Chen, 2015), where the audio datasets showing the strongest annotations in the heat map fell within the 'joy' cluster. Also, another audio study conducted by (Soleymani, et al., 2013) again had a similar heat map result, wherefore the "static annotations" for the contours representing the distribution of annotations once again favoured the 'joy' segment.

Figure 6:12 is a zoomed-in view of Figure 6:11 to include all data points. However, the clip number has been replaced with the name of the film clip.

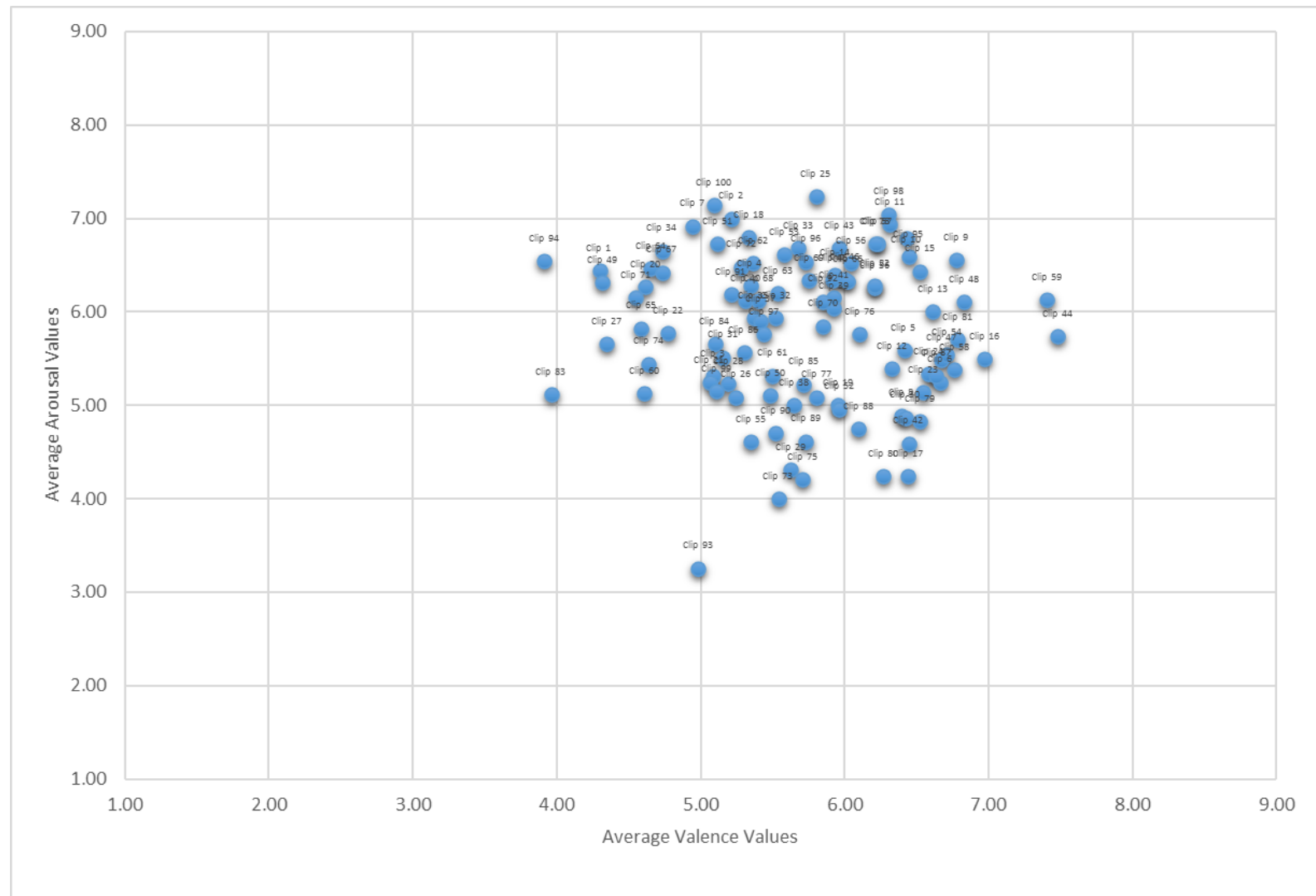
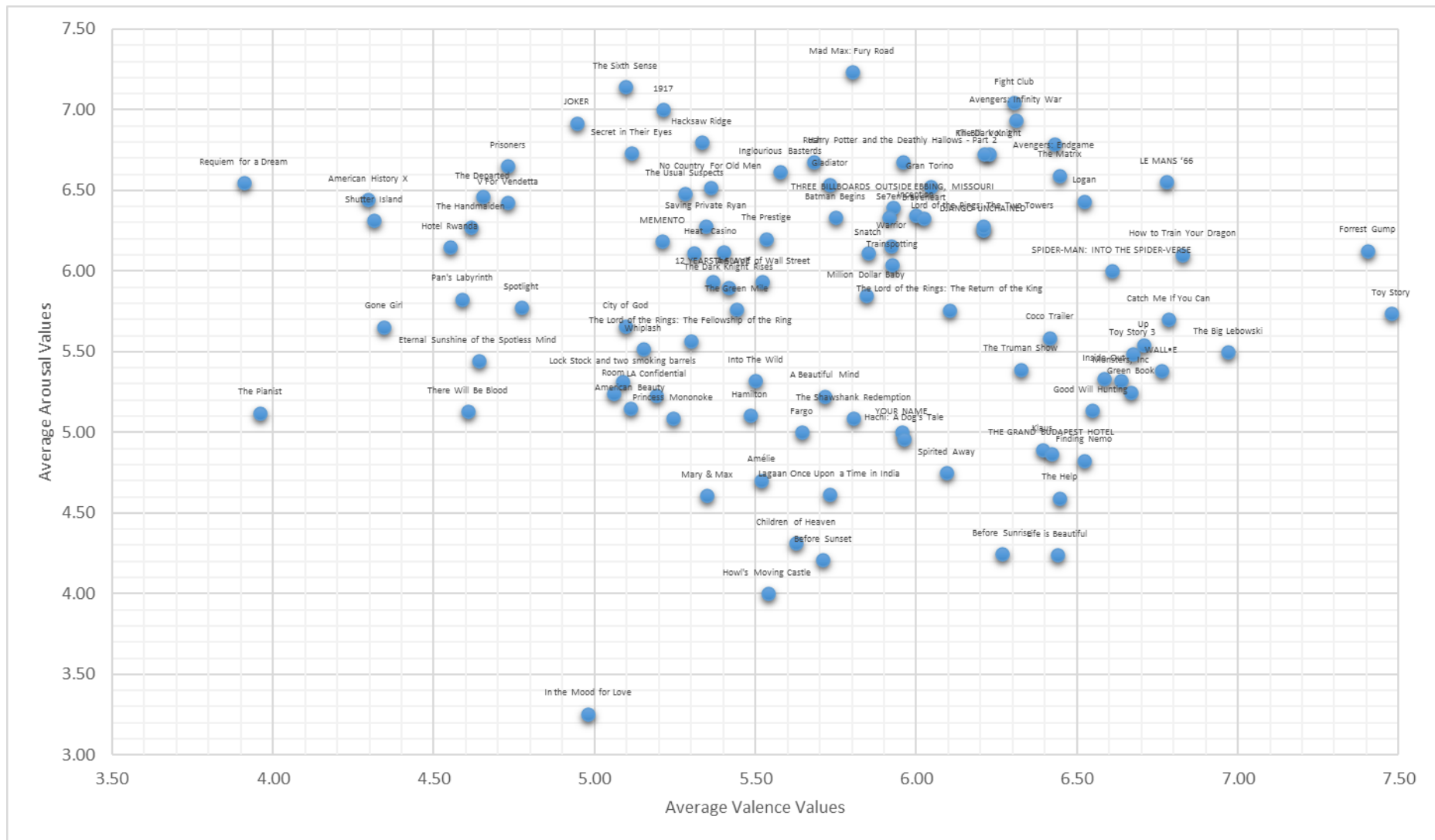


Figure 6:11 - All Clips Average Arousal and Valence on SAM Scale



6.4.4 Comparison to other datasets

This section will evaluate other similar studies before evaluating the performance of this dataset. It is essential to understand previous research approaches in relation related to this study because it helps put its results into context, and highlights the strengths and weaknesses of this study in relation to previous works.

Name of Study	Overview of Study	Size of Study
LIRIS-ACCEDE	Scoring for arousal and valence	9,800 excerpts ranging from 8 to 12 seconds long
EMDB	Ratings for induced arousal, valence, dominance dimensions	52 film clips that were 40 seconds long (no audio)
AVS	Scoring for arousal and valence	100 film trailers ranging from 60 to 226 seconds with original audio

Table 6:2 - Similar studies overview

6.5 Discussion

One of the first discussion points for this research must be how the approaches differ between the similar studies carried out before AVS, and where the AVS fits in relation to this research.

6.5.1 Content length

The LIRIS-ACCEDE model consisted of 9,800 excerpts ranging from 8 to 12 seconds long from 160 featured films with a total runtime of 97,028 seconds (Baveye, et al., 2015); the EMDB consisted of 52 film clips which were 40 seconds long with no audio and a maximum runtime of 2,080 seconds. Finally, the AVS had 100 clips from 60 to 226 seconds with a maximum potential runtime of 143,000 seconds (median of 143-second clips). Source selection here was based on making this study more real-world applicable, which was one of

the main reasons for choosing film trailers according to the criteria listed previously in section 6.3.1.

We can see a significant difference in the average run time of a clip between the two previous studies and the AVS study. So, one of the ways this could have impacted the results is that the emotional reading was taken over a more extended period than its predecessors. It could be argued that film trailers may cover a range of emotions rather than a specific scene, which was utilised in previous research methods.

One study conducted by (Metallinou & Narayanan, 2013) showed that user ratings of their perceived emotions from film clips are not simply averages over the course of the content but are more influenced by highly arousing scenes with low valence.

One of the strengths of this type of content could be that it was limited to film trailers. This is because film trailers are built to generate interest and entice people to watch the film.

6.5.2 Content

The content for LIRIS-ACCEDE was based on 160 movies freely available under a Creative Commons license. This was one of their main goals when creating their affective database, as they believed it should be freely available to the research community. However, the creators reference one of the weaknesses as being low expertise levels with the user-generated content. These movies covered genres that included comedy, animation, action, adventure, thriller, documentary, romance, drama, and horror. Databases of 9,800 excerpts were subsequently extracted from 160 selected movies (Baveye, et al., 2015).

The content of the EMDB includes categories such as erotic, horror, social positive, social negative, scenery, and object manipulation, from which 127 film clips were created whose emotional context was maintained for the entire duration; the clips' content also focused on the presence of human beings. The original 127 film clips were rated by 11 participants using the SAM scale. 52 clips were selected from these ratings, and 75 were discarded due to a high variability from the self-reported ratings (Carvalho, et al., 2012).

6.5.3 Number of Ratings

The number of ratings is one of the other critical components of these datasets. The LIRIS-ACCEDE gathered 582,000 valence annotations from 187,000 comparisons collected from 1,517 annotators and 665,000 arousal annotations from 221,000 comparisons from 2,442 annotators.

6.5.4 Limitations

Some limitations of this study are that every user who did the experiments would be using their own equipment and would be in their own environment, and these are factors we could not control. However, they add value to the study because this is how these systems would operate in real-world environments, as this is where video content will be consumed and recommended to the end user. So, while this would not be the original starting method of this type of research, it did allow for a more real-world setting to gather the results.

6.6 Conclusion

The study presents a new affective video database comprised of 6,202 affective scores across all 100 film clips from over 200 participants. Comparing the outcomes of this study to

the original aims, it is clear that the outcomes have been achieved with the following contributions. Firstly, replicating the strengths of the IAPS while using video as the media has been achieved in the sense that the underpinning foundations for data collection have been replicated successfully from picture to video content.

The subsequent data collected using the above methods are presented in graphs in Figure 6:11 and Figure 6:12 to give an insight into the affective scoring that users provided, allowing us to understand better the role of emotion in movie trailers and video content.

Finally, the dataset is an original dataset with considerable use or replication of currently existing datasets. While it was confined to film trailers, this provided benefits in terms of the duration, the number of scenes, and emotional impact. It is important to note that this was not the potential emotional impact of viewing the whole film, but it did give a much more precise definition of the film trailers' affective scorings.

Additionally, there were some observational contributions from the study, which are as follows:

1. The content potentially caused the tightly packed affective cluster
2. The method seemed to have positive statistical results compared with other studies
3. The data provided a fascinating insight into possible affective space for genres which could be adopted into systems like Netflix, Amazon Prime and YouTube in future for affective content categories.

6.7 Summary

This chapter presented the methods, systems, and initial results for the AVS dataset. The first objective of this chapter was to create a dataset upon which we could test our theories around the effectiveness of an IAPS approach to create a video alternative in the AVS dataset. The analysis of these results was carried out in a similar fashion to the way that they were presented in the original IAPS.

Section 6.3 outlined the methods adopted to conduct this experiment and the tools and mechanisms used in its execution. Finally, in Section 6.4, we reviewed the results from the experiment, providing necessary descriptive statistics and tables in a similar vein to those presented in IAPS studies, as well as looking at the distribution based on an interpretation of the circumplex model in (Yazdani, et al., 2013) and comparing it to other datasets.

6.8 Progression

The results from this study showed similar statistical importance to those presented in the original IAPS, presenting an original dataset in the AVS based on video content and the methods used to create datasets of this nature in the future. The next progression for this research is to extract features from video content and then use regression methods to test this dataset's effectiveness for predicting arousal and valence.

Chapter 7 - Predicting Affective Responses for Video Content Using Machine Learning

7.1 Introduction

This chapter aims to provide a method for affective video content analysis. The affective content analysis could be used as a new way to recommend content to users and as a new way to categorise the content. This could prove helpful in creating a new method for categorising video content based on features that can identify prevailing emotions that users may experience or feel during a given scene. Providing a tool to find out which parts of the movie are, for example, joyful, sad, or angry, and being able to add this type of emotional data to video content, would enhance recommendations to users as it provides an extra dimension to the recommendation process that is currently not considered.

Ultimately, we could move away from a genre-based recommendation and categorisation process to an emotional recommendation and categorisation process, or one where the two processes work side by side.

Another critical aspect of the subjectivity of emotion is that this is a very personal thing, and one user may experience one emotion for a given piece of content. In contrast, another user experiences an entirely different emotion. So even with this approach, there would still need to be a crowdsourced aspect to benchmark how accurate this system is. For example, Netflix is currently utilising the thumbs up or thumbs down rating system to obtain binary input for its recommendation process, which only provides a broad idea of a user's likes or dislikes.

One study that investigated different aspects of the above was (Simon, et al., 1999), who looked at whether there was a consistent and specific relationship between valence and arousal when self-reporting. The study concluded that there was clear evidence that how the stimulus is delivered affects the emotional response.

Also, a study conducted by (Detenber, et al., 1997) investigated the effects of pictures on individual emotional reactions to images. This study concluded that picture motion influences the emotional response to a given image.

7.2 Aims

This study aims to take the AVS dataset created in Chapter 6 and use it to create a model moving towards producing an affective recommendation system using automatic recognition methods outlined in this chapter. Specifically, the chapter will build, train, and test machine learning regression models for the prediction of emotional arousal and valence, and investigate how effective this study is by using standard statistical measures of performance and presenting these results.

The overview of the aims of this chapter is as follows:

- Investigate feature extraction methods that can be utilised on video content
- Devise a range of regression models using machine-learning techniques
- Test these methods using machine learning techniques to gauge their effectiveness

7.3 Methods

The following section provides an overview of the methods used to extract statistical features from video content. As this proved difficult to recreate from pre-existing studies, it was decided to find new approaches and ways of doing this.

7.4 Feature extraction

One of the main research issues with this type of project directly relates to feature selection, which should be used to represent affective scoring correctly. This is due to the lack of knowledge within this field regarding the relationships between features and affect. Therefore, we have tested numerous features to try and draw out as much semantic information as possible, to find out which features may help us solve this research issue.

It was important to review as many features as possible in the literature. Some of these features, or similar ones, have been utilised in other studies and have proven to be effective. However, current research does not provide a consistent or concise feature set for generating affective results in video analysis. Another issue was the lack of transparency in some research to replicate the methods they used to specify features. These obstacles meant it was important to capture as many features as possible so they could be included in the modelling stage.

7.4.1 QCTOOLS (Quality Control Tools for Video Preservation)

QC Tools is a free, open-source software that helps analyse digitised video files via audio-visual analytics and filtering. It is a tool for extracting audio and video features (Coalition, 2021).

7.4.2 RGB (Red, Green, and Blue)

To gather the RGB data for the video clips, the first step was to extract all the frames using a MATLAB script for each of the 100 video clips. This resulted in over 300,000 video frames that had to be analysed and processed. These were then run through a MATLAB script that calculated each film clip's average red, green, and blue average.

7.4.3 Audio toolbox MATLAB

The audio toolbox MATLAB was designed to provide tools for audio processing, speech analysis and acoustic measurement (Mathworks, 2021). The prime use of this toolbox was to extract audio features from the video files into statistical measures so that they could be implemented into machine learning models.

7.4.4 Music Information Retrieval (MIR) Toolbox

The MIRtoolbox was designed to offer a set of functions for MATLAB dedicated to extracting audio and musical features (Jyväskylä, 2021).

7.4.5 MATLAB feature extraction

MATLAB feature extraction is a repository of MATLAB scripts for extracting audio features. (JuliusGruber, 2017).

7.5 Extracted Features

With a combination of the tools covered in Section 7.4, a total of 142 features were extracted.

7.5.1.1 Audio features

The audio features were extracted from several different tools highlighted in Section 7.4.

The features extracted are as follows: A set of 35 features from the Audio toolbox MATLAB were extracted, including energy, energy entropy, fundamental frequency, zero-crossing rate, spectral centroid, spectral centroid (spread), spectral entropy, spectral flux, spectral roll-off, the first 13 MFCCs, harmonic ratio, and 12 chroma vectors. These used means and standard deviations. For each feature, the mean and standard deviation were calculated.

For example, the MIR feature set has 54 features, which is the mean of the following features: dynamics, spectral centroid, brightness, spread, skewness, kurtosis, RollOff95, RollOff85, specEntropy, flatness, roughness, irregularity, MFCCs, deltaMFCC, deltaDeltaMFCC, zero-crossing rate, low energy, spectral flux, and tonal features.

7.5.1.2 Video features

The extracted video features were as follows, utilising the tools covered in Section 7.4. A MATLAB script was written to extract the average RGB from the extracted frames of the video files that provided average R, G, and B. Utilising QC Tools feature extraction for video, the following features were obtained: Y Average (brightness of a picture), Yrang, U Average (chroma planes), V Average (chroma planes), TOUT Average (Temporal Outliers), TOUT count, SAT broadcast Count (Saturation), SAT illegal Count, HUE Average (Hues), BRNG Average (identifies the number of pixels which fall outside the standard video broadcast range), BRNG Count, bit rate, DAR(aspect ratio), Frames/Dur.

7.6 Machine Learning Techniques

The following machine learning techniques were employed to develop regression models in MATLAB using the regression learner. This approach was advantageous, as it allowed for the testing of multiple forms of regression models on the data collected. For example, the following appeared in the top three at least once in the cross-validation 5 or cross-validation 10 in the arousal and valence results.

We decided against breaking out the data into a test set so we could get a clearer idea of the overall results at this stage, as we were testing, and it was necessary to gauge the effectiveness of this method. This was a compromise of practicality to provide the most significant insight into the AVS dataset.

7.6.1 Regression models

This study utilised regression models to investigate the hypothesis between the features extracted and their relationship to arousal and valence data provided by the end-users. Linear regression attempts to model the relationship between variables.

$$y = \beta_0 + \beta_1 x + \varepsilon,$$

Equation 3 - Simple linear regression model (Rencher & Schaalje, 2008)

Equation 3 shows a simple linear regression model with y as the dependent or response variable and x as the independent or predictor variable, β_0 the intercept, and β_1 the regression coefficient (slope). This leaves the random variable ε , which is the error term in this model. Error is a representational statistical term representing random fluctuations,

measurement of errors, or the effect of factors outside of our control (Rencher & Schaalje, 2008).

7.6.2 Cross-validation

Cross-validation refers to a technique that is essential where a sampling procedure is used when evaluating machine learning models and how well they performed for an independent test data set. This study utilised a 5-fold cross-validation and 10-fold cross-validation to try and assess the overall effectiveness of the model and its validity.

In the subsections that follow, the performance benchmarks utilised in this study are described.

7.6.3 Model performance metrics

This section will cover the statistics provided within the MATLAB regression learner for a point of reference.

7.6.3.1 Mean squared error (MSE)

MSE is used to highlight how close the regression line is to assess data points. The smaller the mean square error, the closer the data. Which is shown as the line of best fit (Chicco, et al., 2021). The best value for MSE is zero, because it would mean all data points fit on the line of best fit, but it is not typically possible to achieve this for real data in machine learning evaluations.

Terms for MSE formula:-

MSE = mean squared error

m = number of data points

i = index

X_i = observed values

Y_i = predicted values

$$MSE = \frac{1}{m} \sum_{i=1}^m (X_i - Y_i)^2$$

Equation 4 - MSE Formula (Chicco, et al., 2021)

7.6.3.2 Root mean squared error (RMSE)

The RMSE is the square root of the mean of the square of all the errors and is considered excellent for general error metrics within datasets (Neill & Hashemi, 2018).

Terms for RMSE formula :-

$RMSE$ = root-mean-square error

i = index

n = sample size

S_i = forecasts (expected values or unknown results)

O_i = observed values (known results)

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (S_i - O_i)^2}$$

Equation 5 - RMSE Formula (Neill & Hashemi, 2018)

7.6.3.3 R-squared (R^2)

R^2 is the coefficient of determination that analyses the differences in variables and gives you the percentage of variation in the dependent variable, one explained by the independent variables (Chicco, et al., 2021). For this research, the dependent variables were ground-truth arousal and valence, and the independent variables were the predictions made by the machine learning models.

Terms for R^2 formula :-

R^2	=	coefficient of determination
m	=	number of observations
i	=	index
X_i	=	value of the variable to be predicted
Y_i	=	value in a sample
\bar{Y}	=	mean value of a sample

$$R^2 = 1 - \frac{\sum_{i=1}^m (X_i - Y_i)^2}{\sum_{i=1}^m (\bar{Y} - Y_i)^2}$$

Equation 6 - R² Formula (Chicco, et al., 2021)

7.6.3.4 Mean absolute error (MAE)

MAE is the average of all absolute errors (Chicco, et al., 2021). MAE is considered the most lateral measure of average error size, and it is an unambiguous measure of average error size (Willmott & Matsuura, 2005).

Terms for Mean absolute error formula :

MAE = mean absolute error

m = total number of data points

i = index

X_i = prediction

Y_i = true value

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^m |X_i - Y_i|$$

Equation 7 - MAE Formula (Chicco, et al., 2021)

7.6.3.5 Prediction speed (~obs/sec)

Prediction speed is how quickly the model takes to be built in MATLAB from the presented data. In addition, there may be useful statistics when looking at how efficient the presented statistical measures would be for use in a real-world application, giving an indication of processing times and efficiency.

7.7 Results

This study's results are promising compared to similar studies shown in Figure 7:7. The results based on RMSE do not fall between the typical 0.2 and 0.5 acceptable ratios to be considered a relatively accurate prediction in the conventional sense. However, there is a significant improvement in the results presented for the LIRIS-ACCEDÉ Discrete dataset.

While this leads to the conclusion that there is certainly room for improvement to bring the model into a traditionally acceptable range, it is important to note that it is an emerging technology. Future work on AVS development could be to add more features over time, which may lead to its improvement, as this in theory, could help improve its accuracy.

Some regression models performed better than others, as highlighted in the tables below.

This data was provided in detail to clearly understand exactly where this model was, rather than providing the baseline statistics.

This study found that the following regression methods can be considered most useful for this application of research:

- Exponential GPR
- Cubic SVM
- Rational quadratic GPR
- Bagged trees
- Coarse Gaussian SVM
- Medium Gaussian SVM
- Linear SVM

7.7.1.1 Model performance metrics

Due to the nature of this research domain, it can be hard to directly compare any two studies because there may be subtle differences in how the research was conducted and the presented outcomes. This led to two conclusions, which were that comparisons are not strictly possible, and that it would be best to report multiple statistical performance measurements to give a more accurate insight into the performance of the methods used in the study.

The results presented below are based on the average statistical performance shown in the MATLAB Regression Learner. The following Table 7:1, Table 7:2, Table 7:3 and Table 7:4 have been included due to numerous models having similar performance levels. If this research were to be utilised in the future, this might influence the final choice.

7.7.1.2 The Top 10 Best Performing Models (RMSE) Cross-5 – Arousal

	1	2	3	4	5	6	7	8	9	10
Top 10 Cross 5 – Arousal	Cubic SVM	Exponential GPR	Rational quadratic GPR	Quadratic SVM	Medium Gaussian SVM	Coarse Gaussian SVM	Bagged Trees	Linear SVM	Coarse Tree	Boosted Trees
RMSE (Validation)	0.66	0.66	0.66	0.67	0.68	0.68	0.69	0.69	0.73	0.74
R-squared (Validation)	0.33	0.33	0.33	0.31	0.28	0.28	0.27	0.26	0.18	0.15
MSE (Validation)	0.43	0.44	0.44	0.45	0.46	0.47	0.47	0.48	0.53	0.55
MAE (Validation)	0.48	0.51	0.50	0.49	0.54	0.55	0.53	0.53	0.57	0.60
Prediction speed (~obs/sec)	3000	2400	2500	3100	3000	3200	1700	2200	3100	1400

Table 7:1 - Top 10 Best Performing Models (RMSE) for Cross 5 – Arousal

Table 7:1 shows the top ten best models based on RMSE for the Cross-5 arousal ratings. It highlights the statistical difference between the measurements of the top performer for RMSE; the top-performing value was 0.66 and the eighth 0.69, which shows some consistency in the dataset, as multiple regression methods report similar results. However, when we look at the R-squared for the top-performing model, the value is 0.33, and the eighth is 0.26, which is significantly lower. The same consistency seems to appear for the MSE values and MAE values. This may mean this approach is not limited by a single best-performing method but would require further research and studies to explain why this might be happening. Overall, the Cubic SVM seems to be the best-performing regression method.

7.7.1.3 The Top 10 Best Performing Models (RMSE) Cross-10 – Arousal

	1	2	3	4	5	6	7	8	9	10
Top 10 Cross 10 – Arousal	Bagged Trees	Exponential GPR	Coarse Gaussian SVM	Coarse Tree	Medium Gaussian SVM	Rational Quadratic GPR	Cubic SVM	Quadratic SVM	Boosted Trees	Linear SVM
RMSE (Validation)	0.67	0.68	0.69	0.69	0.69	0.69	0.71	0.73	0.73	0.74
R-squared (Validation)	0.30	0.28	0.27	0.27	0.26	0.25	0.23	0.18	0.18	0.16
MSE (Validation)	0.45	0.47	0.47	0.47	0.48	0.48	0.50	0.53	0.53	0.54
MAE (Validation)	0.52	0.53	0.55	0.52	0.55	0.53	0.52	0.53	0.59	0.57
Prediction speed (~obs/sec)	930	1300	1800	1900	1400	1700	1700	1700	1100	1800

Table 7:2 - Top 10 Best Performing Models (RMSE) for Cross-10 – Arousal

Table 7:2 shows the top ten best models based on RMSE for the Cross-10 arousal ratings. An interesting observation is that when using Cross-10 validation compared to Cross-5, it seems to have a small impact on performance. However, with the Cross-10, we can see a statistically significant trend for the RMSE and the MAE between the top ten results, with the R-squared and MSE changing more significantly with a Cross-10 validation. Once again, across these results, we can see a trend of consistency between the different linear methods. Overall, bagged trees are the best regression method from these results.

7.7.1.4 The Top 10 best performing models (RMSE) Cross-5 – Valence

	1	2	3	4	5	6	7	8	9	10
Top 10 Cross-5 – Valence	Linear SVM	Medium Gaussian SVM	Exponential GPR	Rational Quadratic GPR	Matern 5/2 GPR	Coarse Gaussian SVM	Bagged Trees	Cubic SVM	Quadratic SVM	Squared Exponential GPR
RMSE (Validation)	0.66	0.66	0.67	0.67	0.67	0.69	0.70	0.70	0.71	0.75
R-squared (Validation)	0.22	0.21	0.21	0.20	0.20	0.16	0.13	0.12	0.09	0.00
MSE (Validation)	0.44	0.44	0.44	0.45	0.45	0.47	0.49	0.49	0.51	0.56
MAE (Validation)	0.50	0.53	0.53	0.53	0.53	0.56	0.56	0.56	0.57	0.61
Prediction speed (~obs/sec)	3600	3700	3400	3100	3500	3600	1900	3300	3400	3200

Table 7:3 - Top 10 best performing models (RMSE) for Cross 5 – Valence

Table 7:3 shows the top ten best models based on RMSE for the Cross-5 valence ratings. Once again, we can see a similar pattern for the results for the RMSE, which are comparable to the arousal results. However, R-squared was noticeably lower than in the arousal scores. While it is not suggested that these two scores are directly related, it is an exciting observation that valence seems to be based on a different set of factors than arousal. Overall, linear SVM is the best regression method from these results.

7.7.1.5 The Top 10 best performing models (RMSE) Cross 10 – Valence

	1	2	3	4	5	6	7	8	9	10
Top 10 Cross 10 – Valence	Coarse Gaussian SVM	Medium Gaussian SVM	Exponential GPR	Rational Quadratic GPR	Matern 5/2 GPR	Bagged Trees	Linear SVM	Cubic SVM	Squared Exponential GPR	Fine Gaussian SVM
RMSE (Validation)	0.68	0.68	0.69	0.70	0.70	0.70	0.72	0.74	0.75	0.75
R-squared (Validation)	0.17	0.17	0.16	0.13	0.13	0.13	0.08	0.03	0.00	0.00
MSE (Validation)	0.46	0.47	0.47	0.49	0.49	0.49	0.51	0.55	0.56	0.56
MAE (Validation)	0.55	0.53	0.54	0.54	0.54	0.55	0.57	0.58	0.61	0.61
Prediction speed (~obs/sec)	1800	1900	1600	1800	1700	1100	1900	1900	1700	1900

Table 7:4 - Top 10 best performing models (RMSE) for Cross 10 – Valence

Table 7:4 shows the top ten best models based on RMSE for the Cross-10 valence ratings. It can be observed in the comparisons across five valence results that the Cross-10 validation had slightly lower statistical results. However, these follow similar patterns as observed in the Cross-5 validation. SVM methods and GPR methods still resulted in the top three. There was a more significant drop-off in R-squared within these results. Overall, coarse Gaussian SVM is the best regression method for this research.

7.8 Discussion

After reviewing the above data, it becomes clear there is some causality within the data that is being measured by the different regression methods, producing a similar statistical result.

The following is a condensed table that summarises this chapter.

	5-Fold Cross-Validation—Arousal	10-Fold Cross-Validation - Arousal		5-Fold Cross-Validation - Valence	10-Fold Cross-Validation - Valence
Best performing regression method					
Method	Cubic SVM	Bagged Trees		Linear SVM	Coarse Gaussian SVM
RMSE (Validation)	0.66	0.67		0.66	0.68
R-squared (Validation)	0.33	0.30		0.22	0.17
MSE (Validation)	0.43	0.45		0.44	0.46
MAE (Validation)	0.48	0.52		0.50	0.55
Prediction speed (~obs/sec)	3000	930		3600	1800
Second-best performing regression method					
Method	Exponential GPR	Exponential GPR		Medium Gaussian SVM	Medium Gaussian SVM
RMSE (Validation)	0.66	0.68		0.66	0.68
R-squared (Validation)	0.33	0.28		0.21	0.17
MSE (Validation)	0.44	0.47		0.44	0.47
MAE (Validation)	0.51	0.53		0.53	0.53
Prediction speed (~obs/sec)	2400	1300		3700	1900
Third-best performing regression method					
Method	Rational Quadratic GPR	Coarse Gaussian SVM		Exponential GPR	Exponential GPR
RMSE (Validation)	0.66	0.69		0.67	0.69
R-squared (Validation)	0.33	0.27		0.21	0.16
MSE (Validation)	0.44	0.47		0.44	0.47
MAE (Validation)	0.50	0.55		0.53	0.54
Prediction speed (~obs/sec)	2500	1800		3400	1600

Table 7:5 - Direct comparison of the top 3 regression methods

The information in Table 7:5 shows a clear picture of the regression analysis. We can see that for arousal, and the 5-fold cross-validation cubic support vector machine regression method performed best (RSME = 0.66, R-Squared = 0.33). The line of perfect prediction,

along with the real and predicted values for each clip, can be seen in Figure 7:2. The residuals can be seen in Figure 7:3, and the prediction error can be seen in Figure 7:1.

For valence, we can see that 5-fold cross-validation was the linear support vector machine regression method that performed best (RSME = 0.66, R-Squared = 0.22).

The best-performing regression method for 10-fold cross-validation arousal was bagged trees (RSME = 0.67, R-Squared = 0.30). The line of best fit can be seen in Figure 7:5. The residuals can be seen in Figure 7:6, and the prediction error can be seen in Figure 7:4.

The best-performing model for valence 10-fold cross-validation was coarse Gaussian SVM (RSME = 0.68, R-Squared = 0.17).

There was a noticeable difference between the R-squared values between arousal and valence in the same dataset, which has been noted in other studies in the realm of affective research. There was also a noticeable difference between the R-squared values for the cross-5 validation and a 10-fold cross-validation for arousal and valence. Overall, the best-performing regression methods were based either on SVM or GPR.

7.8.1 Discussion of 5-Fold Cross-Validation Arousal

The results presented in Table 7:5 showed that the 5-fold cross-validation results are the most significant, and further merit discussion around what the data may be telling us.

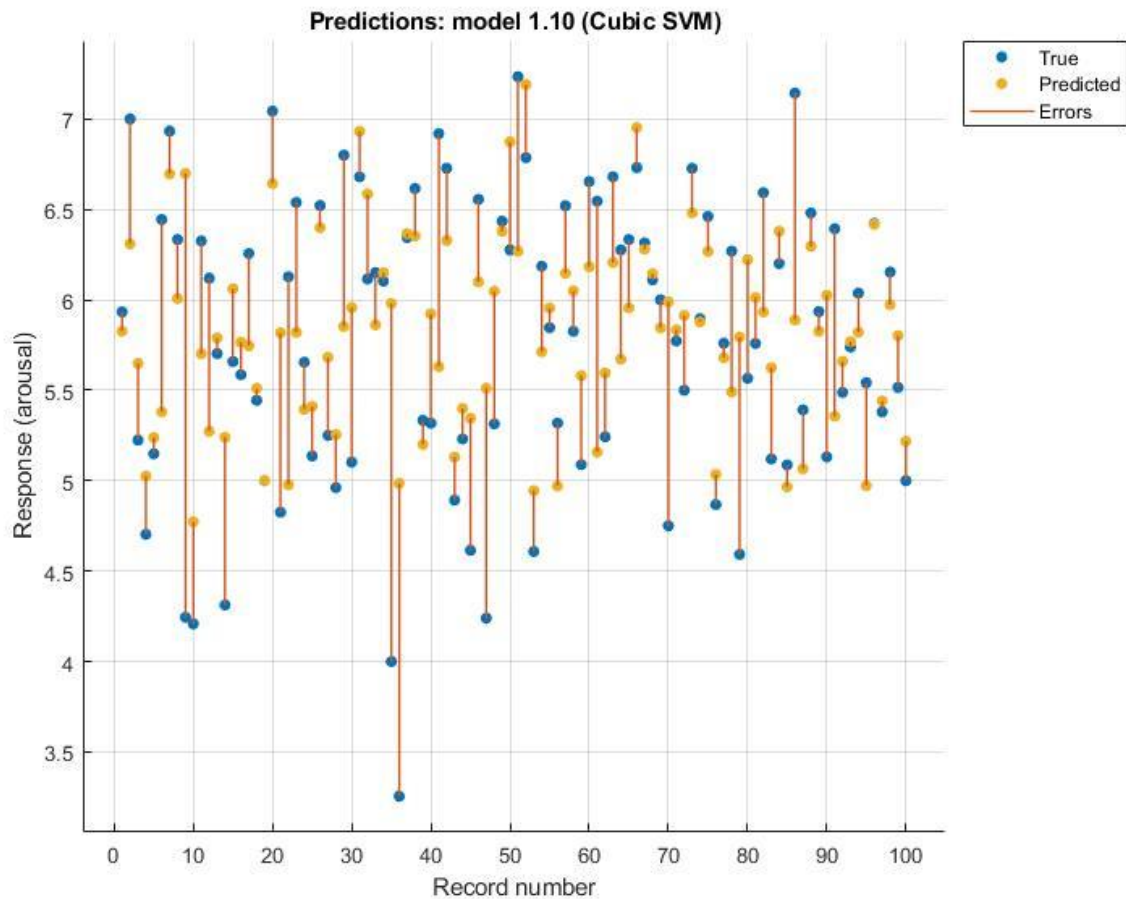


Figure 7:1 - 5-Fold Cross-Validation - Arousal True Vs Predicted (Cubic SVM)

Figure 7:1 shows the predicted values versus the true values for each clip, based on the average arousal results. When reviewing Figure 7:2, it is clear this method would be considered conservative, because most of the points did not fall on the line of a perfect prediction. As we can see, there is some spread in the data between the true and the predicted. This is highlighted by the error bars, as we can see that all the predictions are within 2.5 points of their predictors. For example, highlighted by data point 9 was a true arousal score of 4.2 and a predicted arousal score of 6.7. This resulted in a - 2.5 difference at

the extreme end. Overall, the model tends to predict higher for lower true values and predict lower for higher true values.

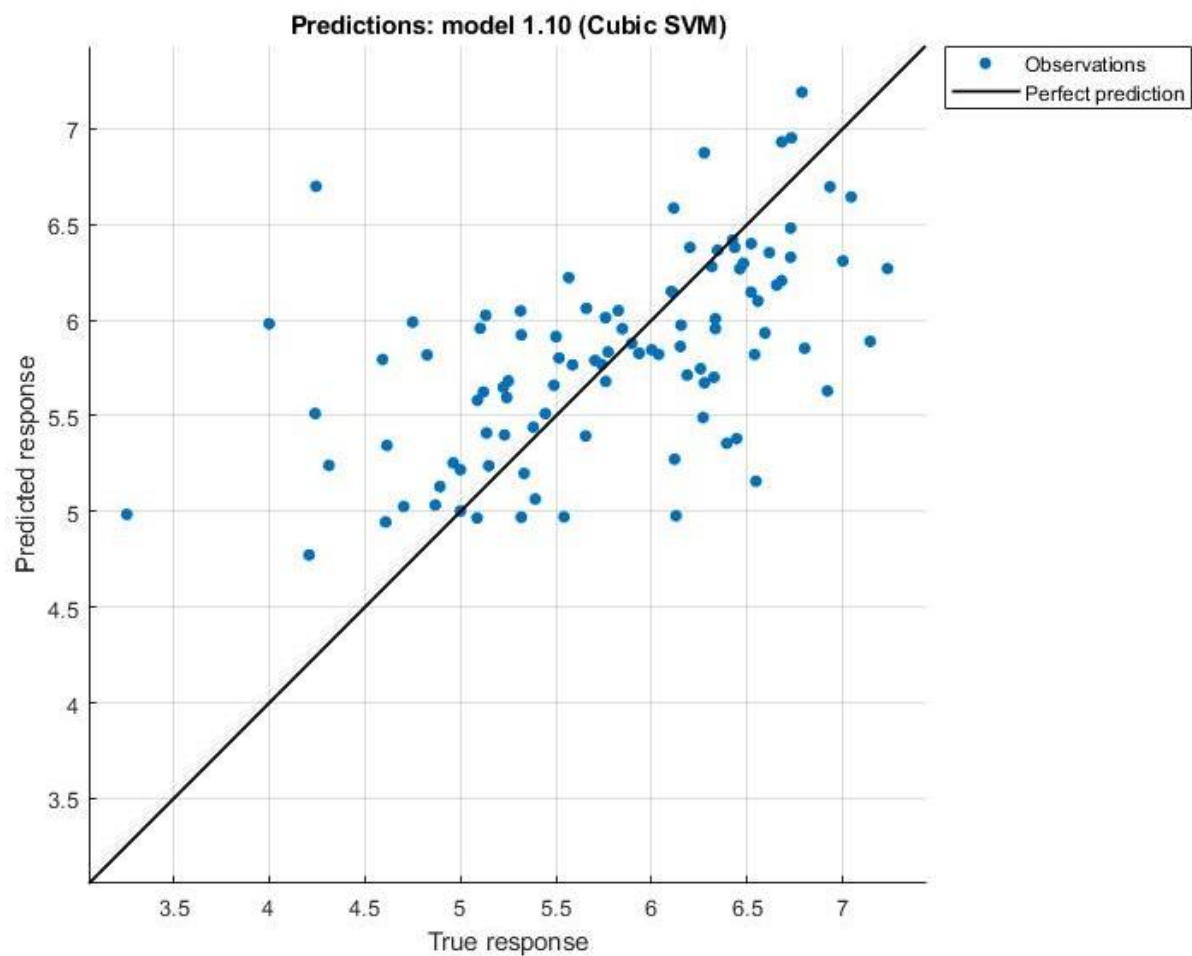


Figure 7:2 - 5-Fold Cross-Validation - Arousal Predicted vs Actual (Cubic SVM)

Figure 7:2 shows the individual clips plotted against the line of perfect prediction, indicating that most of the data points for the true response are between 4 and 7 and that most of the data points for the predicted response fall between 4.5 and 7. This also confirms that the model tends to predict higher for lower true values and lower for higher true values.

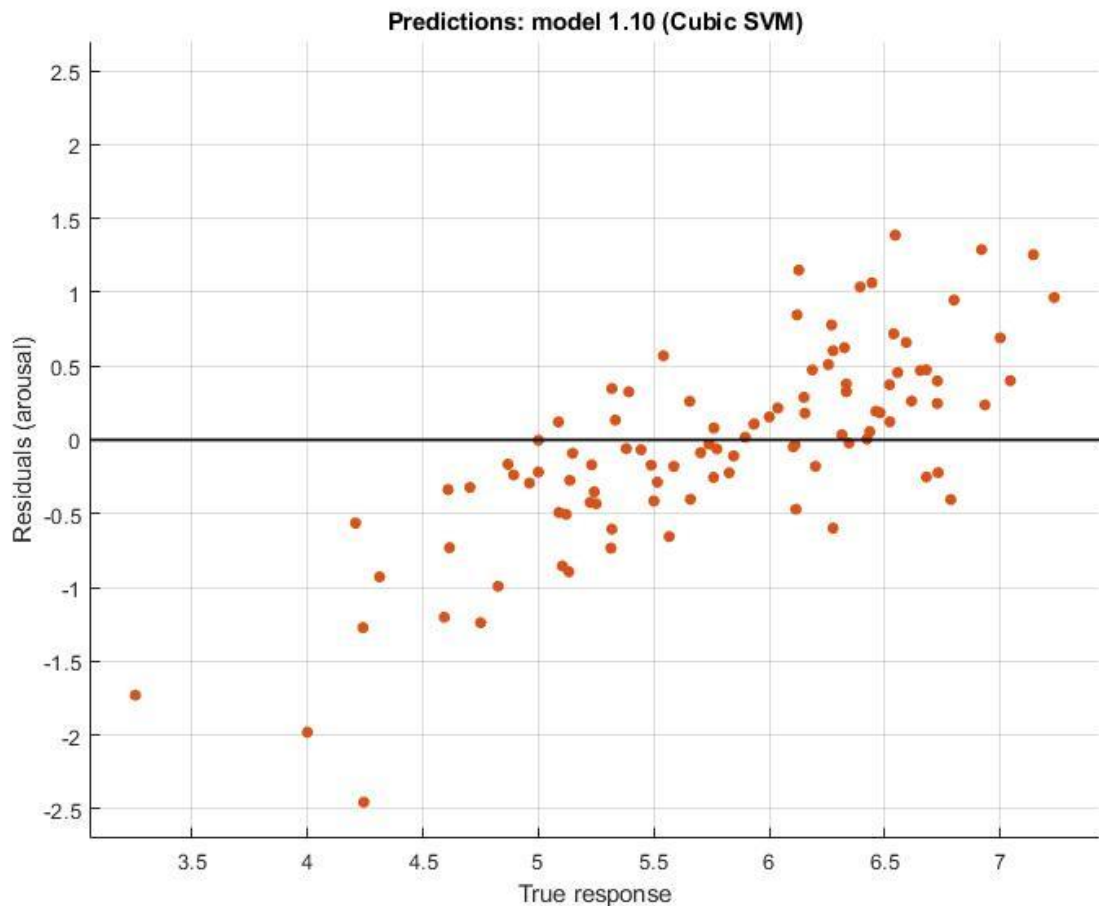


Figure 7:3 - 5-Fold Cross-Validation - Arousal Residuals (Cubic SVM)

Figure 7:3 shows the residual plots where the data largely falls between 1.5 and -1.5 based on the residual plots. However, the results in the context of the 1 to 9 scale that they were scored upon provided a greater context to these results. In simple terms, the movements on the provided scale at one point imply a different emotional result.

With this in mind, Figure 7:3 shows the residuals where we can see six clips have values above 1 and six clips have the value of below -1. This means that only 12% of the clips fell outside what could be considered an acceptable emotional range.

The final consideration is that when you consider the subjective nature of emotions and that 88% of the clips fell under 1 and -1, respectively, the presented results may have more validity to end-users than the statistics currently suggest. However, this claim would require further study and experimentation.

7.8.2 Discussion of 5-Fold Cross-Validation Valence

After reviewing the best-performing model results in Table 7:5 for valence, we can see that a Linear SVM model performed the best statistically.

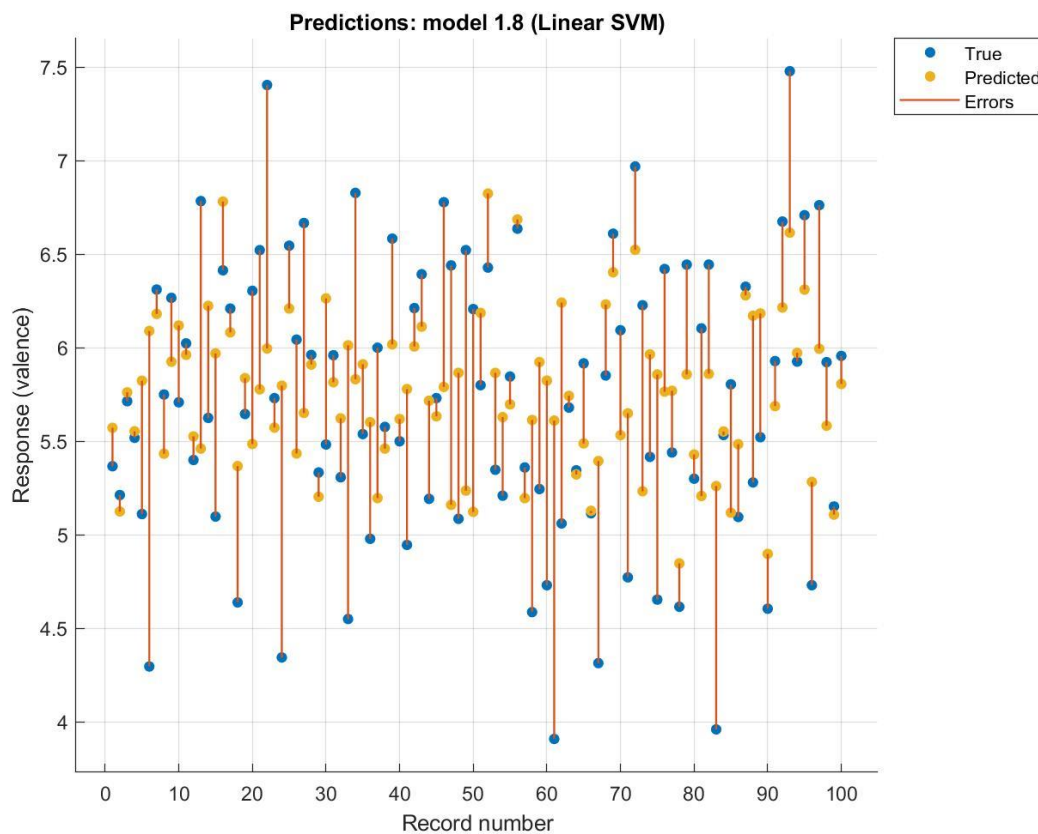


Figure 7:4 – 5-Fold Cross-Validation—Valence True Vs Predicted (Linear SVM)

Figure 7:4 shows the predicted values versus the true values for each clip, based on the average arousal ratings. When reviewing Figure 7:4, it is again clear that this method would be considered 'conservative', as most of the points did not fall on the line of perfect prediction.

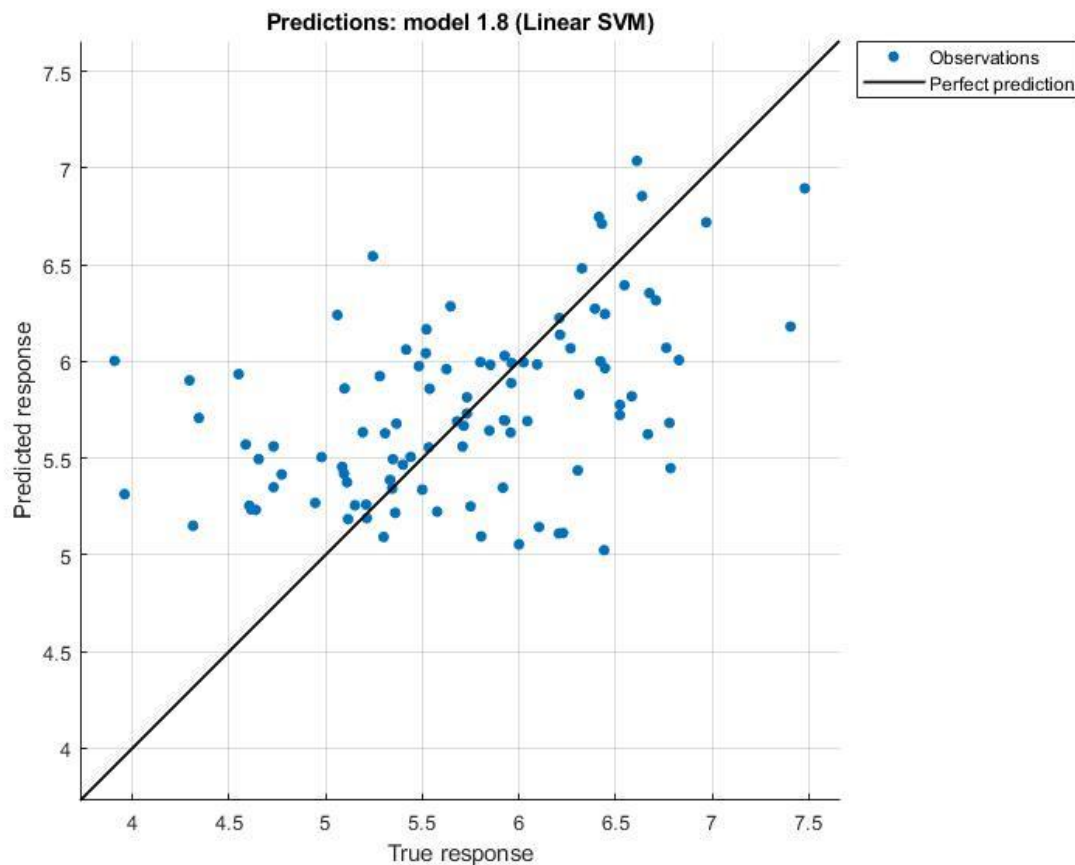


Figure 7:5 - 5-Fold Cross-Validation - Valence Predicted vs Actual (Linear SVM)

Figure 7:5 shows the individual clips plotted against the line of perfect prediction, indicating that most of the data points for the true response are between 4.3 and 6.8 and that most of the data points for the predicted response fall between 5 and 7. This also confirms that the model tends to predict higher for lower true values and lower for higher true values.

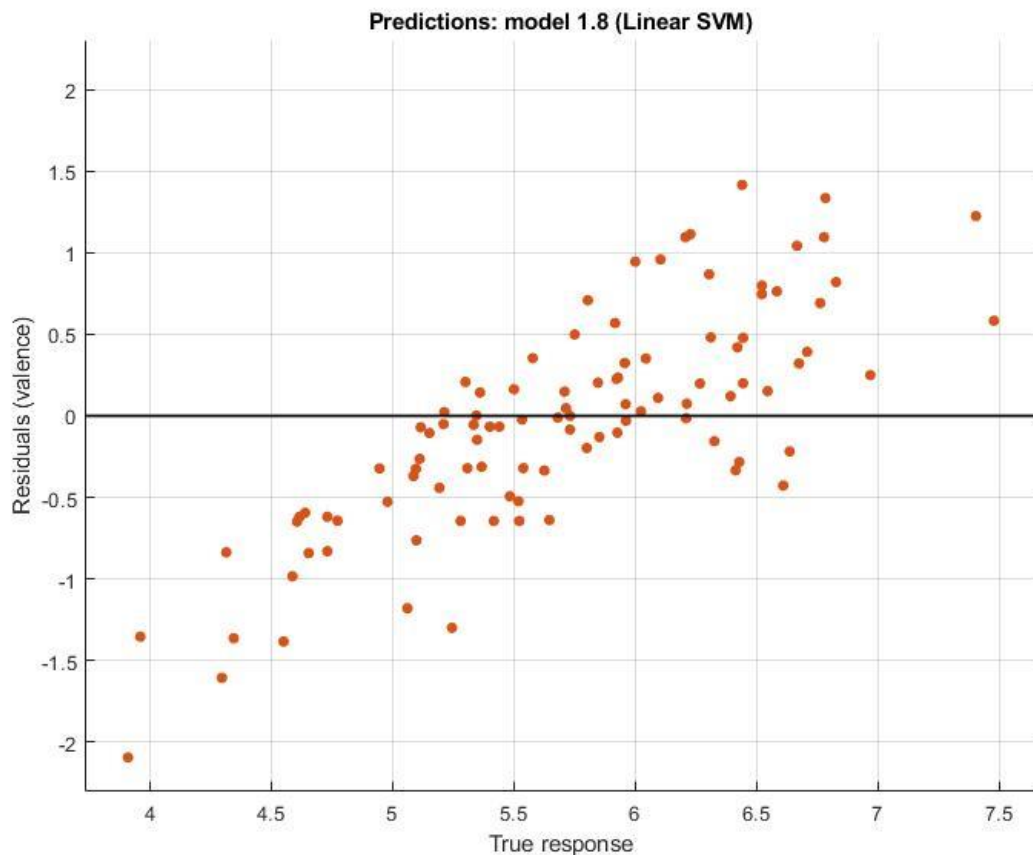


Figure 7:6 – 5 -Fold Cross-Validation - Valence Residuals (Linear SVM)

Figure 7:6 shows the residual plots where the data largely falls between 1.5 and -2.1 based on the residual plots. However, once again, the results in the context of the 1 to 9 scale that they were scored upon provided a greater context to these results. In simple terms, the movements at one point on the provided scale imply a different emotional result.

With this in mind, Figure 7:6 shows the residuals where we can see seven clips have values above 1 and seven clips have the value of below -1. This means that only 14% of the clips fell outside what could be considered an acceptable emotional range.

The final observation is that when you consider the subjective nature of emotions, and that 86% of the clips fell under 1 and -1, respectively, the presented results may have more validity to end-users than the statistics currently suggest. However, this claim would require further study and experimentation.

7.9 Conclusion

In conclusion, this study used the dataset created in Chapter 6 and conducted regression analysis on the dataset to see how effective these methods were for predicting arousal and valence.

From an arousal perspective, the model can remove up to 33% of the variance based on the R-squared values for the top-performing models. However, the valence models can only be considered to account for around 22% of the variance. Principal Component Analysis (PCA) was experimented with but did not improve the performance of the models.

The results from the study provide fascinating insights into how this type of method can be utilised to determine arousal and valence from video content. To compare these findings with a similar study, the closest in the literature would be LIRIS-ACCEDE (Baveye, et al., 2015).

	LIRIS-ACCEDE (Protocol A)		AVS	
Metrics	Arousal	Valence	Arousal	Valence
Pearson's r	0.31	0.31	0.57	0.4
R-squared	0.1	0.1	0.33	0.21
MSE	0.30	0.30	0.44	0.44

Figure 7:7 - AVS compared with LIRIS protocol A results

Figure 7:7 shows the statistical values reported in the presented study and the LIRIS-ACCEDE study for discrete emotions. When comparing the R-squared values for arousal, we can see

that this method significantly outperforms LIRIS-ACCEDE (R-squared 0.1 vs R-squared 0.33 for arousal and R-squared 0.1 vs R-squared 0.21 for valence). This cannot be considered a direct comparison. However, due to differences in the studies (please see Section 6.5), it is the closest dataset currently available in terms of content, method, and analysis.

Overall, this study highlights that affect may require a different method that provides more accurate predictions. However, before drawing this conclusion, it would be worth expanding and re-testing some of the video content, and potentially using this method with a pre-existing dataset to compare their affective annotations with its methods.

Another potential benefit of this study is that these methods could aid filmmakers in identifying whether the intended emotions are conveyed for a given piece of video content. This could allow filmmakers to test whether their scene portrays the correct emotional inference.

7.10 Summary

This chapter presents the methods for testing the AVS dataset with regression. The methods for feature extraction and the tools used to facilitate each set of features were described. This affective testing method is based on the traditional approaches outlined in the IAPS model, which is widely adopted for affective picture scoring. It also incorporates numerous elements at the AVS's core as building blocks for the process that could expand on the IAPS methods, transferring them to video content. These results were then presented and compared to LIRIS-ACCEDE R scores LIRIS-ACCEDE (R-squared 0.1 vs R-

squared 0.33 for arousal and R-squared 0.1 vs R-squared 0.21 for valence) highlighting that this method is statistically a much better approach for video content in the realm of affect.

Section 7.3 discussed the methods utilised in the study and primarily focused on this study's features and feature extraction methods. Finally, Section 7.5 highlighted the extracted features from the techniques covered in the previous chapter to provide insight into the features used to try and benchmark arousal and valence.

Concluding in Section 7.7, the overall model performance could account for over 30% of the variance for arousal and over 20% of the variance for valence, while noting statistical improvements on the reporting of the LIRIS study. However, highlighting regression methods may not be the most appropriate way to solve this research issue, but further study would be required to reach this determination.

Chapter 8 - Conclusion

This thesis investigated methods for determining affective scoring for video content. It focused on the best practices for this task, specifically for film trailer clips. The motivation behind this research was an increased interest in the effects of emotion and cognitive, social and natural processes. This has created the need for reliable emotional identification and classification techniques.

We now have a massive amount of multimedia content that has strained current recommendation system methods. This means that increasingly more resources are being dedicated to improving the recommendation processes, for the systems.

An overview of the thesis is as follows. We started by looking at the recent literature and approaches that had been adopted and are currently in use to inform the next stages of the study. This led to the need to present a framework to provide an overview of how an affective system could work from a theoretical perspective. This led us to look at a pre-existing model for affective studies in IAPS, testing whether this method was transferable to video content. We then investigated the perceivable video features from a user perspective using the repertory grid technique.

The findings from the previous studies led us to build a web-based system to collect affective information based on previous work with the IAPS, adapting it to video content to build the AVS system and dataset. It utilised the SAM scale to capture arousal and valence scores for the video content. Once we had created the AVS affective dataset, we extracted several video and audio features from the video content. These were then compared with

the AVS dataset, using regression models to conclude the research. Showing for arousal that the 5-fold cross-validation cubic support vector machine regression method performed best (RSME = 0.66, R-Squared = 0.33) and that for valence, the 5-fold cross-validation was the linear support vector machine regression method that performed best (RSME = 0.66, R-Squared = 0.22).

8.1 Original research questions

As specified in the introduction chapter of this thesis, the research questions to be addressed, and now with their respective findings, can be presented:

8.1.1 What changes would need to be made to the International Picture System (IAPS) to provide affective scoring for video

Chapter 4 took the above research question, and developed the methods and systems to test it. This resulted in a test where the IAPS content was recreated as closely as possible in video content and then user-tested. The results highlighted that the IAPS could feasibly be adopted for use in the AVS system. This contribution was also utilised in Chapter 6 for the AVS system and data collection.

8.1.2 What video features are most important to the users in their perception of video, and from these features, which can be used in machine learning regression methods?

The next research question addressed the lack of knowledge about what features can be used to determine affective scores and are useful to affective systems. This research question was answered in Chapter 5, where a study was presented using the repertory grid method to collect the most important features from users. This was led with reference to

the constructs provided by the users to features that can be used with regression. The outcome was 'light' and 'dark', which could be measured statistically, and 'white', which can be measured with RGB.

8.1.3 How can a dataset be created for video which can be utilised in the same way as the IAPS?

This research question was answered in Chapter 6 with the creation of the AVS dataset, which then presents a new affective video database comprised of 6,202 affective scores across all 100 film clips from over 200 participants in the same manner as the IAPS dataset. The study provided insights into possible affective space for genres, which could be adopted into systems like Netflix, Amazon Prime, and YouTube in future for affective content categories.

8.1.4 What features can be extracted from video content that can be used to predict affect?

This research question was addressed in Chapter 7 by utilising a combination of QCTOOLS, a MATLAB script to extract RGB, the Audio toolbox MATLAB, the MIRtoolbox and MATLAB Feature Extraction. The tools are covered in section 7.4. These resulted in a total of 142 video and audio features that were extracted.

8.1.5 Which learning regression models will best predict affective scores for video content?

This research question was addressed in Chapter 7 by presenting the top 10 best-performing models based on RMSE for arousal and valence using 5-fold and 10-fold cross-

validation. Concluding for arousal that the 5-fold cross-validation cubic support vector machine method performed best (RSME = 0.66, R-Squared = 0.33) and that for valence, the 5-fold cross-validation was the linear support vector machine method that performed best (RSME = 0.66, R-Squared = 0.22).

8.2 Contributions

The following sections summarise the contributions to knowledge made in this thesis concerning affective computing and its application for video content.

The first contribution of this thesis was the introduction of an adaptive framework for affective studies. One of the first research issues encountered during this thesis was the scale and breadth that affective computing could span, the complex nature of the field, and the need for a multidisciplinary approach. The introduction of the AVS Framework (AVSF) can feasibly be tweaked to fit any requirements or system by building around the core components presented in Chapter 3. Within the confines of this thesis, the framework has helped define the affective system presented in simple terms. This allows for the development of affective inputs such as self-reported data via the online AVS system and affective processing with video feature extraction methods and machine learning methods covered in Chapter 8.

The second contribution of this thesis was a method that modifies the IAPS approach and applies it to video content. This utilised similar video content to the original IAPS images to directly compare with the original IAPS results. This study confirmed that IAPS methods could be transferred to video content. This was important because there were fundamental

differences between video content and picture content, the main difference being that video comprises multiple images across different frames. Typically, video content has audio elements which pictures do not. The audio was removed from the clips used in this study, so it was as close to IAPS as possible. This meant it was necessary to explore whether this method could be transferred, whilst recognising the fundamental differences between the content of the two datasets.

The third contribution of this thesis is a list of important perceptual features from users of film clips produced using the repertory grid methodology. This study shed light on what users perceive as important by utilising a psychological research method that allowed them to express themselves freely. While this study was particularly interested in features that could statistically be benchmarked for use with regression modelling, there were other discernible elements that could be investigated and implemented as an affective process, potentially helping the overall accuracy of this type of system. For example, feature analysis was used to determine which clips show humans, and to investigate the importance of their presence to users.

The fourth was the creation of an affective dataset for film clips in the form of the AVS database. It combined the theory of previous studies into a system that could conduct affective research online, allowing video content to be uploaded and its affective results to be collected from users in a system that could be used in future affective studies and research.

This is a unique dataset, not a reproduction or modification of others. The AVS builds upon pre-existing research methods and presents a new affective video database comprising

6202 arousal and valence scores acquired using the SAM scale across 100 film clips from over 200 participants. With further research and investment, this could become the video equivalent of the IAPS for affective research in video. This contribution addresses the fact that other datasets comprise video content most users would not view in their day-to-day life, highlighting one of the unique aspects of this study, which was to weave real-world elements into rigorous experimental design, producing a system and dataset that filmmakers and streaming services could feasibly use to understand the emotional context of the content, without the need for complex methodologies.

This thesis's fifth contribution evaluates various machine-learning approaches to determine the best-performing regression models. This provided clear statistical insight into the potential relevance for real-world applications, and which of these regression methods was best suited. These results can now be used in affective video systems to present users with affective categorisations based on the methods and systems outlined in this thesis.

Benchmarking the AVS based on the R-squared showed that the best-performing models from an arousal perspective can remove up to 33% of the variance, for the top-performing models. However, the valence models only accounted for around 22% of the variance. We also saw that by comparing the R-squared values for this method, it significantly outperformed LIRIS-ACCEDE (R-squared 0.1 vs R-squared 0.33 for arousal and R-squared 0.1 vs R-squared 0.21 for valence) highlighting that this method is statistically a much better approach for video content in the realm of affect.

This study has ways in which it can be adopted into real-world applications, by presenting methods to extract features from video content, which can then be used in conjunction with machine learning systems. For example, these methods could be used alongside current recommendation and categorisation systems to add affective elements to their decision-making processes.

8.3 Future Research: Affective features

Chapter 7 comprises the methods and tools used to extract several audio and video features. A more in-depth analysis is required of these features to discover which yielded the most significant statistical results when calculating affective scoring. The current system has gauged the overall effectiveness of the features provided, even if it is currently only in a black box way. An expansion of the current research could be firstly expanding the number of features within this model, and secondly doing a detailed analysis of these features to find out which help calculate affective scoring, and which are not. This could also be useful when viewing this from an Explainable Artificial Intelligence (XAI) perspective, as the AVS has presented the groundwork to build upon, towards one of these systems. This system has the potential to become autonomous with its feature extraction methods.

There is also potential in looking at machine learning methods that can account for the temporal aspect of video, which was noted during this thesis but never fully explored.

The AVS methods and dataset have provided a well-utilised pre-existing method in IAPS and successfully transferred this to video content. With further work on its methods and tools, the AVS could potentially be as significant for affective video studies as the IAPS has been for affective picture studies. This future work could start by expanding the current AVS dataset, building in more film clips, and increase the number of participants from which the self-reporting methods are gathered, and adding additional user data and affective scoring. One study that could be conducted would be an in-person study within a lab environment, to compare the lab results against the 'real world' results.

References

- Annika Waern, Z. A. D. S., 2009. *An In-Game Reporting Tool for Pervasive Games*. Athens: ACM.
- Baveye, Y., Dellandrea, E., Chamaret, C. & Chen, a. L., 2015. LIRIS-ACCED: A Video Database for Affective Content Analysis. *IEEE Transactions on Affective Computing, Institute of Electrical and Electronics Engineers*, 1(6), pp. 43-55.
- Bazire, M. a. B. P., 2005. *Understanding context before using it..* Berlin, Springer, pp. 29-40.
- Betella, A. & Verschure, P. F. M. J., 2016. The Affective Slider: A Digital SelfAssessment Scale for the Measurement of. *PLOS ONE*, pp. 1-11.
- Black, N. T. & Ertel, W., 2011. *Introduction to Artificial Intelligence*. London: Springer.
- Bradley, M. a. L. P., 2007. The International Affective Digitized Sounds (; IADS-2): Affective ratings of sounds and instruction manual.. *Tech. Rep.*, Issue 3.
- Bradley, M. M. & Lang, P. J., 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), pp. 49-55.
- Canini, L. B. S. a. L. R., 2012. Affective recommendation of movies based on selected connotative features. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(4), pp. 636-647.

Carvalho, S., Leite, J., Galdo-Álvarez, S. & Gonçalves, Ó. F., 2012. The Emotional Movie Database (EMDB): A Self-Report and Psychophysiological Study. *Applied Psychophysiology and Biofeedback*, 37(4), pp. 279-294.

Castellano, G. et al., 2010. *Body gesture and facial expression analysis for automatic affect recognition*. New York: Oxford University Press.

Chen, G. & Kotz, D., 2005. A Survey of Context-Aware Mobile Computing Research. *Dartmouth Computer Science Technical Report*, pp. 1 - 25.

Chen, Y.-A. Y. Y.-H. W. J.-C. C. H., 2015. THE AMG1608 DATASET FOR MUSIC EMOTION RECOGNITION. *ICASSP 2015 IEEE*, pp. 693 - 697.

Chicco, D., Warrens, M. J. & Jurman, a. G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation Davide Chicco1. *PeerJ Comput. Sci.*, pp. 1 - 24.

Cisco, 2020. 2020 Global Networking Trends Report. *Cisco White Paper*, p. 14.

Coalition, B. A. V., 2021. *QCTools Documentation*. [Online]
Available at: <http://bavc.github.io/qctools/>
[Accessed 18 11 2021].

Commons, C., 2021. *About The Licenses*. [Online]
Available at: <https://creativecommons.org/licenses/>
[Accessed 05 03 2021].

Dan-Glauser, E. S. & S. K. R., 2011. The Geneva affective picture database (GAPED): A new 730-picture database focusing on valence and normative significance.. *Behavioral Research and Method*, 43(2), pp. 468 - 477.

Day, J. & Z. H., 1984. The OSI reference model. *Proceedings of the IEEE*, Volume 71, pp. 1334 - 1340.

Detenber, B. H., Simons, R. F. & Bennett, G. G., 1997. Roll 'em!: The effects of picture motion on emotional responses. *Journal of Broadcasting & Electronic Media*, 42(1), pp. 113 - 127.

Dey, A. K. et al., 2004. a CAPpella: Programming by Demonstration of Context-Aware Applications. *ACM Press Proceedings of the 2004 conference on Human factors in computing*, p. 33–40.

Ekman, P., 1999. *Handbook of Cognition and Emotion*. s.l.:Wiley Publishing.

Ekstrand, M., Riedl, J. & and Konstan, J., 2011. *Collaborative filtering recommender systems*. s.l.:Now Publishers Inc..

Felfernig, A. & Burke, R., 2008. Constraint-based Recommender Systems: Technologies and Research Issues. *ACM*, pp. 1-10.

Felix B. Tan, M. G. H., 2002. THE REPERTORY GRID TECHNIQUE: A METHOD FOR THE STUDY OF COGNITION IN INFORMATION SYSTEMS. *MIS Quarterly*, 1(26), pp. 39-57.

Fleureau, J. G. P. a. O. I., 2013. Affective benchmarking of movies based on the physiological responses of a real audience. *Humaine Association Conference on Affective Computing and Intelligent Interaction*, pp. 73 - 78.

Fleureau, J., Guillotel, P. & Orlac, I., 2013. Affective Benchmarcking of Movies Based on the Physiological Responses of a Real Audience. *IEEE Humaine Association Conference on Affective Computing and Intelligent Interaction*, pp. 73 - 78.

Fransella, F. R. B. a. D. B., 2004. *A manual for repertory grid technique*. s.l.:John Wiley & Sons.

Fredrickson, B. a. K. D., 1993. Duration neglect in retrospective evaluations of affective episodes. *Journal of personality and social psychology*, 65(1), p. 45.

Gomez, P. & D. B., 2010. Cardiovascular patterns associated with appetitive and defensive activation during affective picture viewing. *Psychophysiology*, 47(3), pp. 540-549.

Hsu, F.-H., 1999. IBM's Deep Blue Chess grandmaster chips. *IEEE Micro*, 19(2), pp. 70 - 81.

Huntsinger, G. L. C. a. J. R., 2007. How emotions inform judgment and regulate thought. *Trends in cognitive sciences*, 9(11), pp. 393-399.

IMDb, 2021. *IMDb HomePage*. [Online]

Available at: <https://www.imdb.com/>

[Accessed 05 03 2021].

IMDb, 2021. *Top Rated Movies*. [Online]

Available at: <https://www.imdb.com/chart/top/?sort=us,desc&mode=simple&page=1>

[Accessed 28 October 2021].

Izard, C. D. F. B. B. a. K. W., 1974. The differential emotions scale: A method of measuring the subjective experience of discrete emotions.. *Unpublished manuscript - Vanderbilt University*.

Jain, A. K. & Li, a. S. Z., 2011. *Handbook of face recognition*. Vol. 1 ed. New York: Springer.

Jannach, D., Zanker, M., Ge, M. & Groning, M., 2012. *Recommender Systems in Computer Science and Information Systems – A Landscape of Research*. Vienna, Proceedings of the 13th International Conference on Electronic Commerce and Web Technologies.

JuliusGruber, 2017. *github*. [Online]

Available at: <https://github.com/JuliusGruber/Matlab-FeatureExtraction>

[Accessed 15 05 21].

Jyväskylä, U. o., 2021. *MIRtoolbox*. [Online]

Available at: <https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/materials/mirtoolbox>

[Accessed 18 11 2021].

Kelly, G., 1955. *The psychology of personal constructs*. 2 ed. London: New York: Routledge in association with the Centre for Personal Construct Psychology.

Koukounas, E. & O. R., 2000. Changes in the magnitude of the eyeblink startle response during habituation of sexual arousal.. *Behaviour Research and Therapy*, 38(6), pp. 573-584.

Lang, A. & Dietz, R. B., 1999. *Effective Agents: Effects of Agent Affect on Arousal, Attention, Liking & Learning*. San Francisco, Proceedings of the Third International Cognitive Technology Conference.

Lang, P. B. M. a. C. B., 1997. International affective picture system (IAPS): Technical manual and affective ratings.. *NIMH Center for the Study of Emotion and Attention*, Issue 1, pp. 39 - 58.

Lang, P. J., 2014. *Perspectives on Anger and Emotion: Advances in Social Cognition*,. Volume Vi ed. s.l.:Psychology Press.

M. Narasimha Murty, V. S. D., 2011. *Pattern Recognition: An Algorithmic Approach*. s.l.:Springer Science & Business Media.

M.Bradley, M. & J.Lang, P., 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), pp. 49-55.

Marc Hassenzahl, R. W., 2000. Capturing Design Space From a User Perspective: The Repertory Grid Technique Revisited. *INTERNATIONAL JOURNAL OF HUMAN-COMPUTER INTERACTION*, 12(3&4), pp. 441-459.

Marsden, D. & Littler, D., 2000. Repertory grid technique an interpretive research framework. *European Journal of Marketing*, 34(7), pp. 816 - 834.

Marsden, D. & Littler, D., 2000. Repertory grid technique an interpretive research framework. *European Journal of Marketing*, 34(7), pp. 816 - 834.

Mathworks, 2021. *Audio Toolbox*. [Online]

Available at: <https://uk.mathworks.com/products/audio.html>

[Accessed 18 11 2021].

Matthijs Kwak, K. H. ,. P. M. M. B. A., 2014. The Design Space of Shape-changing Interfaces: A Repertory Grid Study. *Proceedings of the 2014 conference on Designing interactive systems*, pp. 181-190.

McDaniel, B. et al., 2007. Facial Features for Affective State Detection in Learning Environments. *Proceedings of the Annual Meeting of the Cognitive Science Society*, Issue 29, pp. 467-472.

Mehrabian, A. & Russell, J., 1974. *An approach to environmental psychology*. s.l.:MIT Press.

Metallinou, A. & Narayanan, S., 2013. *Annotation and Processing of Continuous Emotional Attributes: Challenges and Opportunities*. s.l., 10th IEEE international conference and workshops on automatic face and gesture recognition (FG), pp. 1 - 8.

Molnar, C., 2020. *Interpretable Machine Learning A Guide for Making Black Box Models Explainable*. s.l.:Lulu.com.

Montgomery, D. C., Peck, E. A. & Vining, G. G., 2013. *Introduction to Linear Regression Analysis*. s.l.:Wiley.

Mower, E., Mataric, M. J. & Narayanan, S., 2011. A Framework for Automatic Human Emotion Classification Using Emotion Profiles. *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING*,.

- Neill, S. P. & Hashemi, M. R., 2018. *Ocean Modelling for Resource Characterization*. s.l.:Academic Press.
- Niese, R., Al-Hamadi, A., Heuer, M. & Matuszewski, B., 2011. Machine Vision based Recognition of Emotions using the Circumplex Model of Affect. *IEEE*, pp. 6424 - 6427.
- Oliveira, E., Martins, P. & Chambel, a. T., 2011. iFelt: Accessing Movies Through Our Emotions. *Proceedings of the 9th European Conference on Interactive TV and Video*, pp. 105-114.
- Orellana-Rodriguez, C., Diaz-Aviles, E. & and Nejdl, W., 2015. *Mining affective context in short films for emotion-aware recommendation*. s.l., Proceedings of the 26th ACM Conference on Hypertext & Social Media.
- Ortony, A. & Turner, T. J., 1990. What's Basic About Basic Emotions?. *Psychological review*, 97(4), pp. 315-331 .
- Osgood, C. E., Suci, G. J. & Tannenbaum, P. H., 1957. *The Measurement of Meaning*. Illinois: University of Illinois Press.
- Pease, A. & Pease, B., 2016. *The Definitive Book of Body Language How to Read Others' Attitudes by Their Gestures*. s.l.:Orion.
- Pfaffenberger, B. & Daley, B., 2004. *Computers in your future*. New Jersey: Peason.
- Pflanzer, R. & McMullen, W., 2000. GALVANIC SKIN RESPONSE & THE POLYGRAPH. *BIOPAC Student Lab*, pp. 1 - 23.

- Philippot, P., 1993. Inducing and assessing differentiated emotion-feeling states in the laboratory. *Cognition and emotion*, 2(7), pp. 171-193.
- Picard, R. W., 2003. Affective computing: challenges. *International Journal of human-Computer Studies*, pp. 55 - 64.
- Picard, R. W. & Healey, J., 1997. Affective wearables. *M.I.T Media Laboratory Perceptual Computing*, Volume 1, pp. 231-240.
- Picard, R. W. & Klein, J., 2002. Computers that recognise and respond to user emotion: theoretical and practical implications. *Elsevier*, pp. 141-169.
- Picard, R. W., Vyzas, E. & Healey, J., 2001. Toward Machine Emotional Intelligence: Analysis of Affective Physiological State. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1 - 24.
- Plus, W., 2021. *WebGrid Plus*. [Online]
Available at: <https://grid.eilab.ca/>
[Accessed 07 09 2021].
- Rencher, A. C. & Schaalje, G. B., 2008. *Linear Models in Statistics*. Germany: Wiley.
- Roeckelein, J., 2006. *Elsevier's dictionary of psychological theories*. s.l.:Elsevier.
- Rogers, B. & Ryals, L., 2007. Using Repertory Grid to access the underlying realities in. *International Journal of Market Research*, pp. 595-612.
- Rottenberg, J. R. R. R. & G. J. J., 2007. Emotion elicitation using films. In: J. A. C. & J. J. B. Allen, ed. New York: Oxford University Press, pp. 9-28.

Russell, J., 1980. A circumplex model of affect. *Journal of personality and social psychology*, 39(6), p. 1161.

Russell, J. A., Bachorowski, J.-A. & Fernandez-Dols, J.-M., 2003. FACIAL AND VOCAL EXPRESSIONS OF EMOTION. *Annu. Rev. Psychol.*, Issue 54, pp. 329-349.

Sacharin, V. S. K. & S. K. R., 2012. Geneva Emotion Wheel Rating Study. *Swiss Center for Affective Sciences*, pp. 1-13.

Sander Koelstra, C. M. M. S. J.-S. L. A. Y. T. E. T. P. A. N. I. P., 2012. DEAP : a Database for Emotion Analysis Using Physiological Signals. *IEEE transactions on affective computing*, 3(1), pp. 18-31.

Schachter, S. a. S. J., 1962. Cognitive, social, and physiological determinants of emotional state. *Psychological review*, 69(5), p. 379.

Schaefer, A., Nils, F., Sanchez, X. & Philippot, P., 2009. Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Psychology Press*, 24(7), pp. 1153 - 1172.

Schaefer, A., Nils, F., Sanchez, X. & Philippot, P., 2010. Assessing the effectiveness of a large database. *Psychology Press*, 24(7), pp. 1153 - 1172.

Scherer, K. R., 2005. What are emotions? And how can they be measured?. *Social science information*, 44(4), pp. 695-729.

Scherer, K. R., Bänziger, T. & Roesch, E., 2010. *A Blueprint for Affective Computing: A Sourcebook and Manual*. First ed. Oxford: Oxford University Press.

Schilit, B., Adams, N. & Roy, W., 1994. Context-Aware Computing Applications. *Mobile Computing Systems and Applications*, pp. 85 - 90.

Simon, R. F., Detenber, B. H. & Reiss, T. M. R. A. J. E., 1999. Emotion processing in three systems: The medium and the message. *Society for Psychophysiological Research*, Volume 36, pp. 619 - 627.

Soleymani, M., Caro, M. N. & Schmidt, E. M., 2013. 1000 Songs for Emotional Analysis of Music. *CrowdMM*, pp. 1 - 6.

Statistics How To, 2020. *Statistics How To*. [Online]

Available at: <https://www.statisticshowto.com/probability-and-statistics/correlation-coefficient-formula/spearman-rank-correlation-definition-calculate/>

[Accessed 02 07 2020].

statisticshowto, 2022. *Correlation Coefficient: Simple Definition, Formula, Easy Steps*.

[Online]

Available at: <https://www.statisticshowto.com/probability-and-statistics/correlation-coefficient-formula/>

[Accessed 02 07 2022].

Stevenson, A., 2010. *Oxford Dictionary of English*. 3 ed. Oxford: Oxford University Press.

Tan, E. S.-H., 1995. Film-induced affect as a witness emotion. *Poetics*, 23(1-2), pp. 7 - 32.

Tellegen, A., 1985. Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report.. p. 681–706.

Titchener, E. B., 1898. The postulates of a structural psychology. *Duke University Press on behalf of Philosophical Review*, 7(5), pp. 449-465.

Tok, S. K. M. D. S. a. C. F., 2010. Evaluation of International Affective Picture System (IAPS) ratings in an athlete population and its relations to personality.. *Personality and Individual Differences*, 5(49), pp. 461-466.

UCLA, 2021. *WHAT DOES CRONBACH'S ALPHA MEAN? | SPSS FAQ*. [Online]

Available at: <https://stats.idre.ucla.edu/spss/fag/what-does-cronbachs-alpha-mean/>

[Accessed 14 11 2021].

Vlachostergiou, A., Caridakis, G. & Kollias, S., 2014. *Investigating Context Awareness of Affective Computing Systems: A Critical Approach*. Évry, Conference: 6th International Conference on Intelligent Human Computer Interaction, IHCI.

Watson, D. C. L. a. T. A., 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales.. *Journal of personality and social psychology*, 54(6), p. 1063.

Willmott, C. J. & Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *CLIMATE RESEARCH*, Volume 30, pp. 79 - 82.

Wood, J. R. & Wood, L. E., 2008. Card Sorting: Current Practices and Beyond. *Journal of Usability Studies*, 4(1), pp. 1-6.

Wundt, W., 1896. *Grundriss der Psychologie*. Engelmann.

- Xu, A. H. a. L.-Q., 2005. Affective video content representation and modeling. *IEEE Transactions on Multimedia*, 7(1), pp. 143-154.
- Yang, W. M. K. N. T. K. N. M. M. G. S. T. M. M., 2018. Affective auditory stimulus database: An expanded version of the International Affective Digitized Sounds (IADS-E). *Behavior Research Methods*, 1(15).
- Yazdani, A., Skodras, E., Fakotakis, N. & Ebrahimi, T., 2013. Multimedia content analysis for emotional characterization of music video clips. *EURASIP Journal on Image and Video Processing*, pp. 1- 10 .
- Yoo, H., Kim, M.-Y. & Kwon, O., 2011. Emotional index measurement method for context-aware service. *Expert Systems with Applications*, 38(1), pp. 785 - 793.
- YouTube, 2020. *YouTube*. [Online]
Available at: [youtube.com](https://www.youtube.com)
[Accessed 15 05 2020].
- Zaltman, G., 2003. *How Customers Think : Essential Insights into the Mind of the Market*. s.l.:Harvard Business Press.
- Zhang, S., Huang, Q., Jiang, S. & Tian, W. G. a. Q., 2010. Affective Visualization and Retrieval for Music Video. *IEEE TRANSACTIONS ON MULTIMEDIA*, 12(6), pp. 510-521.

Appendix A Comparison dataset between IAPS and AVS

	IAPS				AVS			
Image/ Video ID	Valence Mean	Valence SD	Arousal Mean	Arousal SD	Valence Mean	Valence SD	Arousal Mean	Arousal SD
IAPS 1019	3.95	1.96	5.77	1.83	3.88	2.34	5.50	2.04
IAPS 1033	3.87	1.94	6.13	2.15	4.81	1.76	5.54	2.19
IAPS 1090	3.7	1.9	5.88	2.15	4.79	1.76	4.73	2.18
IAPS 1202	3.35	1.77	5.94	2.17	3.63	2.17	6.38	2.28
IAPS 1274	3.17	1.53	5.39	2.39	2.90	1.43	4.73	2.42
IAPS 1302	4.21	1.78	6	1.87	3.13	1.71	5.10	2.10
IAPS 1350	5.25	1.96	4.37	1.76	6.10	1.52	4.21	1.86
IAPS 1460	8.21	1.21	4.31	2.63	7.50	1.44	6.08	2.13
IAPS 1620	7.37	1.56	3.54	2.34	5.83	1.53	4.00	1.89
IAPS 1660	6.49	1.89	4.57	2.39	5.81	1.70	4.63	2.19
IAPS 1720	6.79	1.56	5.32	1.82	6.40	1.53	4.27	2.01
IAPS 1731	7.07	1.58	4.56	2.5	5.63	1.38	4.40	1.82
IAPS 1810	6.52	1.49	4.45	2.11	6.25	1.45	5.25	1.74
IAPS 1908	5.28	1.53	4.88	2.15	5.52	1.74	4.40	2.18
IAPS 1932	3.85	2.11	6.47	2.2	5.69	1.68	5.83	1.94
IAPS 2050	8.2	1.31	4.57	2.53	6.96	1.65	5.42	2.02
IAPS 2206	4.06	1.4	3.71	2.03	4.63	0.96	3.63	1.83
IAPS 2580	5.71	1.41	2.79	1.78	5.15	1.38	3.90	2.13
IAPS 2605	6.26	1.45	5.03	2.16	5.90	1.72	5.27	2.26
IAPS 4535	6.27	1.7	4.95	2.32	4.98	1.54	3.98	2.13
IAPS 5040	5.39	1.11	3.75	1.89	4.35	1.68	4.75	2.21
IAPS 5533	5.31	1.17	3.12	1.92	5.29	1.62	4.00	2.07
IAPS 5626	6.71	2.06	6.1	2.19	6.58	1.41	5.79	1.95
IAPS 5700	7.61	1.46	5.68	2.33	5.94	1.90	4.83	2.36
IAPS 5991	6.55	2.09	4.01	2.44	5.98	1.66	2.52	1.66
IAPS 6410	3.49	2.07	5.89	2.28	4.25	1.62	4.50	2.11
IAPS 7034	4.95	0.87	3.06	1.95	4.46	1.40	2.98	1.84
IAPS 7043	5.17	1.26	3.68	2.09	5.00	1.09	3.90	2.03
IAPS 7058	5.29	1.38	3.98	2.17	5.08	1.30	3.42	2.12
IAPS 7240	6.02	1.93	5.51	2.12	5.04	1.52	4.00	2.17
IAPS 7250	6.62	1.56	4.67	2.15	5.58	1.87	4.73	2.19
IAPS 7325	7.06	1.65	3.55	2.07	5.73	1.88	3.85	1.83
IAPS 7354	5.51	1.67	3.73	2.19	5.40	1.25	4.00	2.14

IAPS 7440	6.32	1.61	4.7	1.96	6.04	1.53	4.35	2.09
IAPS 7470	7.08	1.6	4.64	2.26	6.10	1.88	4.63	2.37
IAPS 7580	7.51	1.6	4.59	2.72	5.81	1.71	3.27	1.92
IAPS 8034	7.06	1.53	6.3	2.16	6.42	1.11	5.96	1.70
IAPS 8060	5.36	2.23	5.31	1.99	5.65	1.51	5.71	2.24
IAPS 8251	6.16	1.68	6.05	2.12	5.75	1.83	4.71	2.18
IAPS 9623	3.04	1.51	6.05	1.88	3.52	1.94	5.35	2.20

Table A – Shows a direct comparison dataset between IAPS and AVS that was constructed from our experiments and upon which the results are constructed.

Appendix B AVS film clips results

Film Clip Name	Average Arousal Values	Average Valence Values	Standard Deviation Arousal Values	Standard Deviation Valence Values
American History X	6.44	4.30	2.10	2.30
1917	7.00	5.21	1.60	1.59
12 Years a Slave	5.93	5.37	1.98	2.33
A Beautiful Mind	5.22	5.71	1.67	1.50
Amélie	4.70	5.52	2.20	1.93
American Beauty	5.15	5.11	1.85	2.10
Avengers: Endgame	6.79	6.43	1.47	2.06
Avengers: Infinity War	6.93	6.31	1.54	2.29
Batman Begins	6.33	5.75	2.44	1.87
Before Sunrise	4.24	6.27	2.11	1.60
Before Sunset	4.21	5.71	2.38	1.97
Braveheart	6.33	6.02	1.70	1.81
Casino	6.12	5.40	1.45	1.87
Catch Me If You Can	5.70	6.78	2.13	2.00
Children of Heaven	4.31	5.63	1.86	1.62
City of God	5.66	5.10	1.88	1.83
Coco Trailer	5.59	6.41	1.84	1.82
Django Unchained	6.26	6.21	2.07	1.92
Eternal Sunshine of the Spotless Mind	5.44	4.64	2.05	2.11
Fargo	5.00	5.65	1.86	1.47
Fight Club	7.04	6.30	1.87	1.64
Finding Nemo	4.83	6.52	1.75	1.93
Forrest Gump	6.13	7.40	1.92	1.80
Gladiator	6.54	5.73	1.79	1.80
Gone Girl	5.66	4.34	1.97	2.02
Good Will Hunting	5.14	6.55	1.98	1.65
Gran Torino	6.52	6.04	1.78	1.77
Green Book	5.25	6.67	1.55	1.21
Hachi: A Dog's Tale	4.96	5.96	2.25	2.68
Hacksaw Ridge	6.80	5.33	1.79	2.28
Hamilton	5.10	5.48	2.18	2.18
Harry Potter and the Deathly Hallows: Part 2	6.68	5.96	2.08	1.51
Heat	6.12	5.31	1.53	1.76
Hotel Rwanda	6.15	4.55	1.73	1.88

How to Train Your Dragon	6.10	6.83	1.59	1.89
Howl's Moving Castle	4.00	5.54	2.12	2.02
In the Mood for Love	3.26	4.98	2.17	1.64
Inception	6.34	6.00	1.67	1.85
Inglourious Basterds	6.62	5.58	1.55	2.39
Inside Out	5.33	6.58	1.93	1.91
Into the Wild	5.32	5.50	2.21	1.90
Joker	6.92	4.95	1.77	2.48
Kill Bill: Vol. 1	6.73	6.21	2.25	2.12
Klaus	4.89	6.39	2.33	1.59
LA Confidential	5.23	5.19	1.82	1.36
Lagaan Once Upon a Time in India	4.62	5.73	2.14	2.03
Le Mans '66	6.56	6.78	1.53	1.31
Life is Beautiful	4.24	6.44	1.85	1.85
Lock Stock and two Smoking Barrels	5.31	5.09	2.29	1.82
Logan	6.43	6.52	1.67	1.56
Lord of the Rings: The Two Towers	6.28	6.21	2.17	2.09
Mad Max: Fury Road	7.23	5.80	1.65	2.07
Mary & Max	4.61	5.35	1.53	1.75
Memento	6.19	5.21	1.65	1.68
Million Dollar Baby	5.85	5.85	2.11	1.78
Monsters, Inc	5.32	6.64	1.73	1.89
No Country for Old Men	6.52	5.36	2.02	2.10
Pan's Labyrinth	5.83	4.59	2.10	1.47
Princess Mononoke	5.09	5.24	2.36	2.15
Prisoners	6.65	4.73	1.83	2.52
Requiem for a Dream	6.55	3.91	1.84	1.85
Room	5.24	5.06	2.19	2.33
Rush	6.68	5.68	1.82	1.70
Saving Private Ryan	6.28	5.34	1.71	2.29
Se7en	6.33	5.92	1.69	1.84
Secret in Their Eyes	6.73	5.12	1.87	2.34
Shutter Island	6.31	4.31	2.04	2.18
Snatch	6.11	5.85	2.22	1.73
Spider-Man: Into the Spider-Verse	6.00	6.61	2.36	2.23
Spirited Away	4.75	6.09	2.46	2.07
Spotlight	5.77	4.77	1.82	2.07
The Big Lebowski	5.50	6.97	1.95	1.67

The Dark Knight	6.73	6.23	2.12	2.11
The Dark Knight Rises	5.90	5.42	1.74	1.66
The Departed	6.46	4.65	2.25	1.47
The Grand Budapest Hotel	4.87	6.42	1.83	2.15
The Green Mile	5.76	5.44	2.03	2.52
The Handmaiden	6.27	4.62	2.03	2.19
The Help	4.59	6.44	2.06	1.89
The Lord of the Rings: The Fellowship of the Ring	5.57	5.30	2.34	2.07
The Lord of the Rings: The Return of the King	5.76	6.10	2.17	1.78
The Matrix	6.59	6.44	1.82	2.12
The Pianist	5.12	3.96	1.96	1.77
The Prestige	6.20	5.53	1.75	1.70
The Shawshank Redemption	5.09	5.80	1.93	1.76
The Sixth Sense	7.14	5.10	1.42	2.19
The Truman Show	5.39	6.33	1.83	1.49
The Usual Suspects	6.48	5.28	1.64	1.65
The Wolf of Wall Street	5.93	5.52	1.70	1.81
There Will Be Blood	5.13	4.61	2.36	1.98
Three Billboards Outside Ebbing, Missouri	6.39	5.93	1.89	2.36
Toy Story	5.74	7.48	2.18	1.86
Toy Story 3	5.49	6.67	2.20	2.22
Trainspotting	6.04	5.93	2.01	1.71
Up	5.54	6.71	1.96	1.68
V For Vendetta	6.42	4.73	2.21	1.85
WALL•E	5.38	6.76	1.99	1.84
Warrior	6.15	5.92	2.05	1.65
Whiplash	5.52	5.15	2.25	1.70
Your Name	5.00	5.96	1.73	1.52

Table B - The affective results for the AVS presented in average arousal and valence and standard deviation for the arousal and valence