



TÉCNICO
LISBOA

UNIVERSIDADE DE LISBOA
INSTITUTO SUPERIOR TÉCNICO



Physiological-based Group Emotion Recognition
Novel Methods and Real-World Applications

Patrícia Justo Bota

Supervisor: Doctor Ana Luísa Nobre Fred

Co-Supervisors: Doctor Hugo Humberto Plácido da Silva
Doctor Pablo Santiago César Garcia

Thesis approved in public session to obtain the PhD degree in
Biomedical Engineering

Jury final classification: Pass with Distinction and Honour

2024



TÉCNICO
LISBOA

UNIVERSIDADE DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

Physiological-based Group Emotion Recognition
Novel Methods and Real-World Applications

Patrícia Justo Bota

Supervisor: Doctor Ana Luísa Nobre Fred
Co-Supervisors: Doctor Hugo Humberto Plácido da Silva
Doctor Pablo Santiago César Garcia

Thesis approved in public session to obtain the PhD degree in
Biomedical Engineering

Jury final classification: Pass with Distinction and Honour

Jury

Chairperson: Doctor João Miguel Raposo Sanches, Instituto Superior Técnico, Universidade de Lisboa

Members of the Committee:

Doctor Anna Maria Maddalena Bianchi, Department of Electronics, Information and Bioengineering, Politecnico di Milano, Itália

Doctor Paulo Luís Serras Lobato Correia, Instituto Superior Técnico, Universidade de Lisboa

Doctor Rita Isabel Mangerico Canaipa, Faculdade de Ciências da Saúde e Enfermagem, Universidade Católica Portuguesa

Doctor Sandra Pereira Gama, Instituto Superior Técnico, Universidade de Lisboa

Doctor Hugo Humberto Plácido da Silva, Instituto Superior Técnico, Universidade de Lisboa

Doctor Susana Manuela Martinho dos Santos Baía Brás, Departamento de Eletrónica, Telecomunicações e Informática, Universidade de Aveiro

Funding Institutions - FCT: Fundação para a Ciência e a Tecnologia and
Instituto de Telecomunicações

2024

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my supervisors Prof. Ana Fred, Prof. Hugo Silva and Prof. Pablo Cesar. Namely, for allowing me the opportunity to join their team and continue to learn about machine learning, emotion and physiological data. I would like to thank their positive words, excellent insights, guidance throughout this work, and their prompt availability.

Secondly, I would like to thank Chen Wang and Xinha Net FMCI for the bases to quickly start this project and for their views of the field of emotion recognition, namely the promise of group emotion recognition, which has been instrumental in shaping the direction and success of this work.

Thirdly, my colleagues and supervisors at Centrum Wiskunde & Informatica (CWI), namely Abdallah El Ali and Tianyi Zhang for their expertise and mentorship in the field of emotion recognition and physiological data. I would also like to thank the CWI for the opportunity to work in such a great environment.

Next, I would like to acknowledge my colleagues and friends at IST-IT, Mariana Abreu, Ana Sofia Carmo, Sofia Margarida Monteiro, Rafael Silva, Joana Brito, Vicente Garção, André Gomes, Rui Maciel, Frederico Santos, Afonso Ferreira and many more for making it easy for me to go to work every day.

Lastly, I would like to thank my family and friends for their support throughout this journey, who provided me with everything necessary to make this work possible.

Proofreading of this document has been assisted by large language models.

This work was funded by Fundação para a Ciência e a Tecnologia (FCT) under grants 2020.06675.BD and FCT (PCIF/SSO/0163/2019 SafeFire), FCT/Ministério da Ciência, Tecnologia e Ensino Superior (MCTES) national funds, co-funded EU (UIDB/50008/2020 NICE-HOME), Xinhua Net FMCI (S-0003-LX-18), Ministry of Economy and Competitiveness of the Spanish Government co-funded by European Regional Development Fund (ERDF) (TIN2017-85409-P PhysComp), and Instituto de Telecomunicações (IT), by the Fundo Europeu de Desenvolvimento Regional (FEDER) through the Operational Competitiveness and Internationalization Programme (COMPETE 2020), and by National Funds through the FCT under the LISBOA-01-0247-FEDER-069918 “CardioLeather” and LISBOA-1-0247-FEDER-113480 “EpilFootSense”.

Abstract

Emotions determine human thinking and behaviour. Thus, affective computing is extremely relevant in several applications, from mental health and the creation of personalized services to entertainment.

This work begins by exploring the state of the art in the area of intra-personal emotion recognition through physiological signals, making a quantitative analysis of the various existent approaches.

Next, an interpersonal model was explored that incorporates group dynamics using the WGS methodology, where the synchrony between the physiological signals of a group was studied. This model presented superior results to existing methods and identified relevant synchrony measures. However, the generalization of group analysis was limited by a lack of databases.

To overcome this limitation, a hardware and software infrastructure was developed for the acquisition of physiological signals in groups — the EmotiphAI platform, which includes wearable devices, a collection centre, and a user interface, enabling the simultaneous collection of data from 10 devices at 60 Hz.

Subsequently, a retrospective annotation system was incorporated that selects relevant physiological segments, which demonstrated high accuracy and user satisfaction. This functionality paves the way for the annotation of longer-duration content for the creation of naturalistic datasets.

Using EmotiphAI, the G-REx dataset was created with annotated physiological signals from 190 subjects across 31 movie sessions. Moreover, the respective protocol validated the use of EmotiphAI in a real-world environment.

Overall, this thesis contributes to the advancement of applications in the area of affective computing in the real world through the development of: interpersonal algorithms for emotion recognition; tools for group physiological data acquisition and annotation; and a naturalistic dataset.

Keywords: Emotion Recognition, Physiological Signals, Machine Learning, Group Emotions, Affective Computing

Resumo

As emoções determinam o pensamento e o comportamento humano. Assim, a computação afetiva tem extrema relevância em diversas aplicações, desde a saúde mental e criação de serviços personalizados até ao entretenimento.

Este trabalho começa por explorar o estado da arte na área de reconhecimento de emoções intrapessoais através de sinais fisiológicos, fazendo uma análise quantitativa das abordagens existentes.

De seguida, explorou-se um modelo interpessoal que incorpora a dinâmica de grupo pela metodologia WGS, onde se estudou a sincronia entre os sinais fisiológicos de um grupo. Este modelo apresentou resultados superiores aos métodos existentes e identificou medidas de sincronia relevantes. Contudo, a generalização da análise em grupo estava limitada por falta de bases de dados.

Para colmatar essa limitação, foi desenvolvida uma infraestrutura incluindo programas e o equipamento para a aquisição de sinais fisiológicos em grupo — a plataforma EmotiphAI, que inclui dispositivos vestíveis, um centro de recolha, e uma interface para o utilizador, possibilitando a recolha simultânea de dados de 10 dispositivos a 60 Hz e 20 dispositivos a 60 Hz.

Posteriormente, foi incorporado um sistema de anotação retrospectiva que seleciona segmentos fisiológicos relevantes, que demonstrou elevada precisão e satisfação por parte do utilizador. Esta funcionalidade abre caminho para a anotação de conteúdos de maior duração, mais semelhantes ao mundo real.

Utilizando o EmotiphAI, foi criado o conjunto de dados G-REx com sinais fisiológicos anotados de 190 sujeitos em 31 sessões de cinema. Adicionalmente, o respectivo protocolo validou a utilização do EmotiphAI para contextos do mundo real.

No geral, esta tese contribui para o avanço de aplicações na área da computação afectiva no mundo real através do desenvolvimento: de algoritmos interpessoais para o reconhecimento de emoção; ferramentas para aquisição e anotação de dados fisiológicos em grupo; e um conjunto de dados adquirido num contexto do mundo real.

Keywords: Reconhecimento de Emoções, Sinais Fisiológicos, Aprendizagem automática, Emoções de Grupo, Computação Afetiva

Contents

List of Acronyms	xvii
1 Introduction	1
1.1 Motivation	1
1.2 Objectives	3
1.3 Contributions	5
1.3.1 International Journals	5
1.3.2 International Conferences	7
1.3.3 Student Guidance and Supervision	7
1.3.4 Dissemination Activities	9
1.4 Outline	11
2 Background	13
2.1 Emotion Theories	14
2.1.1 Basic Emotion Theory	14
2.1.2 Appraisal Theories of Emotion	15
2.1.3 Psychological Construction Theory of Emotion	16
2.1.4 Multicomponent Theory of Emotion	17
2.1.5 Discussion	19
2.2 Measuring Emotions	20
2.2.1 Subjective Experience	20
2.2.2 Central Nervous System	22
2.2.3 Behaviour	24
2.2.4 Autonomic Nervous System	26
2.2.5 Discussion	32
2.3 Datasets	33
2.4 Group Emotions	35
2.4.1 Groups	36
2.4.2 Collective Emotions	36

2.4.3	Emotion Transmission in Groups	38
2.4.4	Discussion	40
3	Benchmarking Emotion Recognition	41
3.1	Introduction	42
3.2	Background	43
3.2.1	State of The Art	43
3.2.2	Machine Learning	54
3.3	Methods	55
3.3.1	Data	55
3.3.2	Signal Pre-Processing	56
3.3.3	Feature Extraction and Selection	56
3.3.4	Classification	57
3.3.5	Evaluation	57
3.4	Results	58
3.4.1	Single Modality Models	58
3.4.2	Multi-Modality Models	59
3.5	Discussion	63
3.6	Conclusion	63
4	Group Emotion Recognition	67
4.1	Introduction	68
4.2	Background	69
4.2.1	Group Emotion Classification	69
4.2.2	Metrics for Physiological Synchrony	70
4.2.3	Physiological Datasets for Group Emotion Recognition	71
4.3	Methods	72
4.3.1	Workflow	73
4.3.2	Synchronisation Metrics	74
4.3.3	Data Representation	75
4.3.4	Classification Models	77
4.4	Results	80
4.4.1	Intrapersonal Model	80
4.4.2	Interpersonal Model	83
4.5	Discussion	85
4.5.1	Synchronisation Metrics and Data Representations	85

4.5.2	Interpersonal Model vs Intrapersonal Model	86
4.6	Conclusion	87
5	Data Acquisition	89
5.1	Introduction	90
5.2	Background	91
5.3	Methods	92
5.3.1	Wearable	92
5.3.2	Collector	95
5.3.3	User Interface	97
5.3.4	Technical Validation	98
5.4	Results	99
5.4.1	Device Performance Test	99
5.4.2	Infrastructure Test	103
5.4.3	Sampling Period Test	104
5.5	Discussion	105
5.6	Conclusion	106
6	Emotion Annotation	109
6.1	Introduction	110
6.2	Background	111
6.2.1	Data Annotation	111
6.2.2	Processes Involved in Emotional Self-Report	113
6.3	Methods	113
6.3.1	Annotation Interface	114
6.3.2	Content Segmentation Method	115
Random	115
Scene	116
Electrodermal Activity	116
6.3.3	Data Storage	116
6.3.4	Experimental Study Design	117
6.4	Results	118
6.4.1	Usability Test	119
Statistics	119
Usability	120
Mental Workload	120

6.4.2	Inter-Subject Agreement	120
6.4.3	Self-Reports Coherence	122
6.4.4	Comparison to Reference Annotations	123
6.4.5	Comparison to Electrodermal Activity	124
6.5	Discussion	126
6.5.1	Usability Test	126
6.5.2	Validity Test	126
6.5.3	Content Segmentation Method	128
6.6	Conclusion	129
7	Group Emotion Dataset	131
7.1	Introduction	132
7.2	Background	133
7.3	Methods	133
7.3.1	Dataset Design	134
7.3.2	Data Contents	139
7.3.3	Dataset Summary	143
7.3.4	Technical Validation	143
7.4	Results	146
7.4.1	Physiological Signals	146
7.4.2	Data Characterisation	148
7.4.3	Statistical Analysis	151
7.5	Discussion	152
7.6	Conclusion	156
8	Conclusions and Future Work	159
8.1	Revisiting the Research Questions	159
8.2	Discussion	166
8.2.1	Input Data	166
8.2.2	Emotion Annotation	166
8.2.3	Emotion Classification	167
8.2.4	Moving Forward	167
8.2.5	Affective Computing Applications	168
8.3	Future Work	169
	Appendices	171
	Appendix A Benchmarking Emotion Recognition	171

Appendix B Group Emotion Recognition	176
Appendix C Data Acquisition	183

List of Tables

2.1	List of predominant basic emotions in the literature.	15
2.2	Emotion components, their measuring system and reported dimensions.	20
2.3	Summary of brain regions identified across emotion theories.	23
2.4	Voice correlates for basic emotions.	26
2.5	Typical features extracted from the EDA signal.	29
2.6	Reported Electrodermal Activity (EDA) modal responses across discrete emotions.	30
2.7	Cardiovascular modal responses across discrete emotions.	32
2.8	Literature datasets for emotion recognition using physiological data.	34
2.9	Group examples.	36
3.1	Summary of the state-of-the-art research studies in the field of affect recognition across its main characteristics.	44
3.2	Public datasets explored for emotion classification.	55
3.3	Hyper-parameters and respective search space across classifiers.	57
3.4	Single-modality experimental results best performing supervised learning classifier per dataset, sensor modality and emotion dimension.	58
3.5	Features selected for the arousal dimension in the single-modality methods.	60
3.6	Features selected for the valence dimension in the single-modality methods.	61
3.7	Classifier used in the multi-modal feature fusion method.	62
3.8	Features used in the multi-modal feature fusion method.	62
3.9	Emotion classification single-modality experimental results.	65
3.10	Experimental results for the feature fusion and decision fusion methodologies.	65
4.1	Characerisation of the AMIGOS and K-EmoCon datasets class labels.	72
4.2	Features extracted from the EDA and RR-interval signals.	76
4.3	Synchrony metrics applied for each data space in the interpersonal methodology.	77
4.4	Intrapersonal approach results for the AMIGOS dataset.	81
4.5	Intrapersonal approach results for the K-EmoCon dataset.	82
4.6	Best performing data representation and synchronization metrics for the interpersonal methodology.	84

5.1	Devices described in the emotion recognition literature for EDA and Photoplethysmography (PPG) data acquisition.	93
5.2	R-IoT Default OSC Message Components.	94
5.3	Main Characteristics of the Plux EDA Sensor.	94
5.4	Main characteristics of the PPG PulseSensor.	94
5.5	Metrics analysed in the EmotiphAI Data Acquisition technical validation.	100
5.6	Results for the Data Acquisition platform.	101
6.1	State of the art emotion annotation platforms.	112
6.2	Demographic information from EmotiphAI Annotator validation tests.	118
6.3	EmotiphAI Annotator Usability Questionnaire results.	120
6.4	EmotiphAI Annotator evaluation statistics.	121
6.5	EmotiphAI Annotator inter-subjects agreement results.	122
6.6	Similarity between the subjects' average self-reports for the segmentation methods.	123
6.7	Comparison between the EmotiphAI Annotator and the LIRIS-ACCEDE annotations.	123
6.8	EmotiphAI Annotator methods' arousal annotations correlation to the Mean Affective Profile (MAP).	125
6.9	Comparison between the state of the art and EmotiphAI Annotator best results on the MAP and arousal annotations correlation.	126
7.1	Overview of EmotiphAI Wearable characteristics.	137
7.2	Description of G-REx self-reported emotion annotation scales.	138
7.3	Summary of the G-REx dataset data.	143
7.4	Data quality metrics deployed for analysing the EDA and PPG signal quality in the G-REx dataset.	145
7.5	EDA data quality metrics analysed in the G-REx dataset.	149
7.6	PPG data quality metrics analysed in the G-REx dataset.	149
7.7	Statistical metrics extracted from the G-REx annotations.	151
7.8	G-REx statistic tests.	152
A .1	Features extracted from each sensor modality.	171
B 1	Features extracted from EDA and RR-interval signals.	176
B 2	Results for the Weighted Group Synchrony (WGS) on the AMIGOS dataset for the Arousal dimension.	179
B 3	Results for the WGS for the AMIGOS dataset, Valence dimension.	180
B 4	Results for the WGS for the KEmoCon dataset, Arousal dimension.	180
B 5	Results for the WGS for the K-EmoCon dataset, Valence dimension.	181

B 6 Hyperparameters space values used in the WGS and intrapersonal approach. 182

C 1 Technical specifications for the Zigbee, Bluetooth and WiFi protocols. 183

C 2 Technical specifications of the TP-Link Wireless N 450Mbps and TP-Link MR3020 3G/Wi-Fi routers, used in the EmotiphAI Collector validation tests. 183

List of Figures

1.1	Thesis content overview.	11
2.1	Literature theories on emotion.	19
2.2	Self-Assessment Manikin (SAM) scale for arousal and valence.	21
2.3	Brain regions correlated to basic emotions.	23
2.4	Example of seven upper face action units and how they can be combined to form diverse facial expressions.	25
2.5	EDA signal decomposition.	29
2.6	Prototypical EDA waveform and extracted features.	29
2.7	Prototypical PPG waveform.	31
2.8	Illustration on individual, group-based, and collective emotions.	37
3.1	Histogram of the number of publications surveyed per year of publication.	43
3.2	Histogram with the number of publications per elicitation material and number of sub- jects explored.	52
3.3	Histogram with the number of publications across physiological signals and classifiers.	53
3.4	Typical emotion recognition data classification processing pipeline.	56
4.1	Workflow of the tested methodologies.	73
4.2	Illustration of the data representations: Morphology and Image-based.	77
4.3	1-Dimensional Residual Temporal and Channel Attention Network (RTCAN-1D).	79
4.4	Residual Neural Network (ResNet)-18 architecture.	79
5.1	EmotiphAI Data Acquisition platform and its components.	92
5.2	EmotiphAI R-IoT Wearable device.	95
5.3	EmotiphAI platform workflow.	96
5.4	Example of EmotiphAI's Data Acquisition HDF5 file.	96
5.5	EmotiphAI Data Acquisition user interface.	97
5.6	Routers used in the EmotihAI Data Acquisition technical validation tests.	99
5.7	Obtained loss for the EmotiphAI Data Acquisition Device Performance Test.	103
5.8	Obtained loss for the EmotiphAI Data Acquisition platform Infrastructure test.	104

5.9	EmotiphAI deployments (part 1).	107
5.10	EmotiphAI deployments (part 2).	107
6.1	Content segmentation for retrospective emotion annotation.	114
6.2	EmotiphAI Annotator end-user interface.	115
6.3	EmotiphAI Annotator content segmentation methods for retrospective annotation.	115
6.4	EmotiphAI Annotator HDF5 storage file.	117
6.5	Experimental protocol deployed to collect the data for the EmotiphAI Annotator technical validation.	118
6.6	Example of the temporal alignment of the EmotiphAI Annotator methods' data by timestamps.	119
6.7	EmotiphAI Annotator average annotations for the three segmentation methods.	123
6.8	Example comparing the LIRIS-ACCEDE ground-truth to the EmotiphAI Annotator.	124
6.9	Examples comparing the collected MAP and the average arousal self-reports from the EmotiphAI Annotator.	125
7.1	Photos taken during the G-REx data collection movie sessions.	134
7.2	EmotiphAI platform setup used in the Diferencial cinema sessions.	136
7.3	Experimental protocol for the collection of the G-REx dataset.	138
7.4	G-REx data structure.	140
7.5	G-REx dataset pickle data structure with respective keys.	142
7.6	Histogram with the number of annotations per user and samples collected per movie genre.	150
7.7	Distribution of personalities and age range.	150
7.8	Histogram of the self-reported G-REx annotations.	151
7.9	Data distribution for the mean standardized EDA and normalized Heart Rate (HR) across the arousal self-report scores.	153
7.10	Data distribution for the standardized mean EDA and normalized HR across the valence self-reported scores.	153
7.11	Data distribution for standardized mean EDA data per subject and normalized mean HR per subject across the movie genres.	154
7.12	Data distribution for the arousal and valence self-reports across the movie genres.	154
1	Summary of the features selected in the feature selection step, combining all the datasets.	175

List of Acronyms

ACC	Accelerometer
ADC	Analog-to-Digital Converter
ANOVA	Analysis of Variance
ANS	Autonomic Nervous System
bpm	beats per minute
BVP	Blood Volume Pulse
CNN	Convolutional Neural Network
CNS	Central Nervous System
CV	Cross-Validation
CWI	Centrum Wiskunde & Informatica
DEAP	Dataset for Emotion Analysis using Physiological Signals
DTW	Dynamic Time Warping
EDA	Electrodermal Activity
EDL	Electrodermal Level
EDR	Electrodermal Response
ECG	Electrocardiography
EEG	Electroencephalography
EESD	Eight-Emotion Sentics Data
EMG	Electromyography
ERDF	European Regional Development Fund

ESMAE	Escola Superior de Música, Artes e Espectáculo
EQF	European Qualifications Framework
FCT	Fundação para a Ciência e a Tecnologia
FEDER	Fundo Europeu de Desenvolvimento Regional
fMRI	Functional Magnetic Resonance Imaging
FIR	Finite Impulse Response
GYR	Gyroscope
GSR	Galvanic Skin Response
Hb	Hemoglobin
HR	Heart Rate
HRV	Heart Rate Variability
IMU	Inertial Measurement Unit
IST	Instituto Superior Técnico
IT	Instituto de Telecomunicações
ITMDER	IT Multimodal Dataset for Emotion Recognition
JRC	Joint Research Centre
K-NN	K-Nearest Neighbors
LOSO	Leave-One-Subject-Out
LSTM	Long Short-Term Memory
MAHNOB	Multimodal dataset for Affect Recognition and Implicit Tagging
MAG	Magnetometer
MAP	Mean Affective Profile
MCTES	Ministério da Ciência, Tecnologia e Ensino Superior
M-F1	Macro-F1 Score
NASA-TLX	NASA Task-Load Index
Obj	Objective

OSC	Open Sound Control
PNS	Parasympathetic Nervous System
PET	Positron Emission Tomography
PPG	Photoplethysmography
ReLU	Rectified Linear Unit
ResNet	Residual Neural Network
RQ	Research Question
RSSI	Received Signal Strength Indicator
RTCAN-1D	1-Dimensional Residual Temporal and Channel Attention Network
SAM	Self-Assessment Manikin
SNR	Signal-to-Noise Ratio
SNS	Sympathetic Nervous System
SRJ	SCImago Journal Ranking
STD	Standard Deviation
SUS	System Usability Scale
SVM	Support Vector Machine
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
USA	United States of America
WESAD	Multimodal Dataset for Wearable Stress and Affect Detection
W-F1	Weighted-F1 Score
WGS	Weighted Group Synchrony

Chapter 1

Introduction

This thesis begins with the motivation for this work and articulates the main objectives identified from the existing challenges in group emotion recognition through physiological signals. The chapter then details the contributions of this work, providing an overview of the main publications and outputs. Finally, the chapter outlines the document structure, providing an overview of the content of the thesis.

Contents

1.1	Motivation	1
1.2	Objectives	3
1.3	Contributions	5
1.3.1	International Journals	5
1.3.2	International Conferences	7
1.3.3	Student Guidance and Supervision	7
1.3.4	Dissemination Activities	9
1.4	Outline	11

1.1 Motivation

Emotions are intricately intertwined with our cognition and behaviour, thus influencing a vast array of our daily interactions and decision-making [1]. This makes emotions a powerful phenomenon which the study and understanding, is not just an academic pursuit, but a practical one that can have profound implications for society. Emotion recognition systems have the potential to be used for example to enhance mental health care by providing new ways to monitor and manage well-being, tailor educational software that adapts its speed and content to the learner’s emotional state, or facilitate the creation of personalized content (e.g. games, music, tv shows) directed to the user preferences. These advance-

ments can both improve quality of life, paving the way for innovations that could reshape industries and improve customer experiences.

The scientific field devoted to the exploration of these technologies is nowadays known as affective computing, defined by Picard [2] as *"computing that relates to, arises from, or deliberately influences emotion or other affective phenomena"*. Affective computing involves the development of systems that can learn about human affective states, defined as the subjective experience of feelings, moods and affective traits [3].

The pursuit of emotion recognition systems can be approached in various ways. The theory on emotion reports that emotions can be described as event-driven multi-componential responses [1, 4]. The components of the multicomponent theory typically include a change in 1) the subjective experience (typically assessed by self-reports); 2) behaviour (typically assessed by facial, body, and vocal expressions); 3) central physiology (typically assessed by image and specialised sensor techniques); and 4) peripheral physiology (typically assessed by the autonomic nervous system sensors). All of these components can be used to infer emotional states.

The traditional approach is the use of self-reports, where individuals provide subjective assessments of their emotional state. However, these are often limited by the subject being able to identify their own emotions, and their willingness to provide accurate information, becoming a burden and exhausting task to the subject. A second alternative is the use of body behaviour, where visual and auditory cues such as facial expressions, gestures, and vocal intonations are used to infer emotional states [5]. However, these methods can sometimes be deceived by the discrepancies between expressed and felt emotions. A third approach is the analysis of the Central Nervous System (CNS), either through neuroimaging or sensors such as Electroencephalography (EEG) [5]. These methods are often limited by the need for expensive equipment and for the subject to be in a laboratory setting. Lastly, physiological signals can be the most direct and unfiltered view of emotional states. Signals such as EDA and PPG can be captured unobtrusively via wearable technology, making them suitable for a near-continuous collection of emotion-related data in the real world for the development of emotion recognition systems.

Humans are social beings, such that in our daily lives, we are often surrounded by friends, family or strangers. During these interactions, emotional transfers such as emotional contagion or collective effervescence between the group members modulate the individuals' behaviour and emotions, creating a macro-level emotion denoted as collective emotions. Moreover, the literature reports that information on group dynamics can introduce context information for emotion recognition, with improvement in emotion classification. However, analysing the literature we observe that most datasets collect data in the lab data in individual sessions. Often using small video clips as an elicitation method, not providing a full build-up to emotion nor allowing the development of group dynamics.

Moreover, collecting data in the lab has resulted in the collection of data which has been shown to

differ from the one collected in the real world, either by showing stereotyped responses or by being of low intensity [6]. This has led to poor performance of current datasets and emotion recognition algorithms, which the performance does not generalize to real-life applications.

This gap arises from the absence of tools capable of facilitating the reliable collection of physiological data for groups and emotion annotation tools for real-world settings. Therefore, there is a need for a validated infrastructure that allows the collection of group physiological data, the creation of the respective datasets, and the development of algorithms that can take advantage of this multi-modal data information.

This thesis aims to address these end-to-end limitations from the development of physiological-based data acquisition and emotion annotation tools to the collection of naturalistic group data for the creation of a novel dataset, to the use of group data for the development of emotion classification algorithms, comparing them to intrasubject methods.

1.2 Objectives

The overarching goal of this thesis is to explore group emotion recognition based on physiological signals. This includes developing novel methods and the required underlying tools. To achieve this goal, the following objectives and subsequent Research Questions (RQs) are proposed:

1. **Objective (Obj) 1. Development of Affective Computing Algorithms:** Creation of emotion recognition algorithms through physiological signals focused towards group settings. This involves identifying suitable input physiological signals, the metrics to extract from each sensor, the classification algorithm, the evaluation metrics, and analyzing how the group data can add information to the classification, among others.

RQ 1.1: *What feature set/machine learning algorithm should be applied for emotion recognition?*

RQ 1.2: *What performance can be achieved by predominant datasets in the literature?*

RQ 1.3: *What is the best method to deal with multi-modal data for emotion classification?*

RQ 1.4: *What synchronization metrics and data representations are most suitable for measuring physiological synchrony for emotion recognition?*

RQ 1.5: *Does the emotion classification accuracy improve with the inclusion of group-level information?*

2. **Obj 2. Group-based Physiological Data Collection:** Develop an infrastructure that allows the collection of physiological data in real-world scenarios, both in individual and group data settings.

This involves gathering a hardware base with the physiological sensors, the device form factor, the software for synchronised data collection across multiple devices, and the interface for real-time data visualisation.

RQ 2.1: *How does the sampling rate affect the data loss and transfer quality in multi-device scenarios?*

RQ 2.2: *To what extent does the network infrastructure influence the maximum number of devices that can collect data without data loss?*

3. **Obj 3. Emotion Annotation for Naturalistic Settings:** Design an emotion annotation method that allows the efficient annotation of emotion-related data in real-world scenarios, both from individuals and groups. This involves considering the limitations of naturalistic data collection, the development of the software for the annotation of physiological data, the synchronization of the annotations across multiple devices and sessions, and the creation of the end-user interface for emotion annotation.

RQ 3.1: *Are retrospective annotations in long-duration content usable for emotion annotation?*

RQ 3.2: *How do retrospective annotations compare to conventional approaches in long-duration content?*

RQ 3.3: *Which content segmentation method is more suitable for emotion annotation in long-duration content?*

4. **Obj 4. Real-world Affective Computing Dataset:** Collect a real-world affective computing dataset. This involves the collection of physiological data in a naturalistic scenario, the annotation of the physiological data, proposing a structure for the dataset, and the validation of the data.

RQ 4.1: *Can large amounts of annotated physiological data be collected reliably in a naturalistic setting?*

RQ 4.2: *How does an infrastructure designed for group physiological data acquisition perform in a real-world setting?*

This thesis aims to advance the field of affective computing by addressing critical challenges identified in the area of group emotion recognition using physiological signals, from the validation of group dynamics measured through physiological signals, the development of the infrastructure required for group data collection and emotion annotation of naturalistic data, to the collection of real-world data capable of creating generalisable emotion recognition algorithms.

By fulfilling these objectives, the aim is to contribute to the theoretical and practical understanding of affective computing, ultimately paving the way for innovative applications that can be integrated

into the real world across various domains such as healthcare, education, entertainment, and beyond. Additionally, by exploring the use of group data for emotion recognition, the goal is to provide a more comprehensive understanding of the role of social context in emotion recognition, potentially leading to more context-aware and accurate computing solutions.

1.3 Contributions

This thesis has resulted in 11 publications in international journals (9 in Q1 ranking) and 4 publications in international conferences.

The following listings detail the main works, organized by how they address the main objectives of this work. When available, journal publications include the quartile according to the SCImago Journal Ranking (SRJ), while conference publications include the conference rank according to CORE¹.

1.3.1 International Journals

1. P. Bota, T. Zhang, A. El Ali, A. Fred, H. P. da Silva, and P. Cesar, "Group synchrony for emotion recognition using physiological signals," *IEEE Trans. on Affective Computing*, vol. 14, no. 4, pp. 2614–2625, 2023
SJR: Q1; IF: 9.6
Obj 1. Development of Affective Computing Algorithms
2. P. Bota, R. Silva, C. Carreiras, A. Fred, and H. P. da Silva, "Biosppy: A python toolbox for physiological signal processing," *SoftwareX*, vol. 26, p. 101712, 2024
SJR: Q2; IF: 2.4
Obj 1. Development of Affective Computing Algorithms
3. P. Bota, C. Wang, A. L. N. Fred, and H. Plácido da Silva, "A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals," *IEEE Access*, vol. 7, no. 1, pp. 140990–141020, 2019
SJR: Q1; IF: 3.9
Obj 1. Development of Affective Computing Algorithms
4. P. Bota, C. Wang, A. Fred, and H. Silva, "Emotion assessment using feature fusion and decision fusion classification based on physiological data: Are we there yet?," *Sensors*, vol. 20, no. 17, p. 4723, 2020

¹<https://portal.core.edu.au/conf-ranks/>; Accessed on 27/03/2024.

SJR: Q2; IF: 3.576

Obj 1. Development of Affective Computing Algorithms

5. P. Bota, A. Fred, J. Valente, C. Wang, and H. P. da Silva, "A dissimilarity-based approach to automatic classification of biosignal modalities," *Applied Soft Computing*, vol. 115, p. 108203, 2022

SJR: Q1; IF: 8.7

Obj 1. Development of Affective Computing Algorithms

6. P. Bota, E. Flety, H. P. d. Silva, and A. Fred, "EmotiphAI: a biocybernetic engine for real-time biosignals acquisition in a collective setting," *Neural Computing and Applications*, vol. 35, no. 8, pp. 5721–5736, 2023

SJR: Q1; IF: 4.5

Obj 2. Group-based Physiological Data Collection

7. R. Silva, G. Salvador, P. Bota, A. Fred, and H. Plácido da Silva, "Impact of sampling rate and interpolation on photoplethysmography and electrodermal activity signals' waveform morphology and feature extraction," *Neural Computing and Applications*, vol. 35, no. 8, pp. 5661–5677, 2023

SJR: Q1; IF: 4.5

Obj 2. Group-based Physiological Data Collection

8. M. N. Supelnic, A. F. Ferreira, P. Bota, L. Brás-Rosário, and H. Plácido da Silva, "Benchmarking of sensor configurations and measurement sites for out-of-the-lab photoplethysmography," *Sensors*, vol. 24, no. 1, 2024

SJR: Q1; IF: 3.4

Obj 2. Group-based Physiological Data Collection

9. P. Bota, P. Cesar, A. Fred, and H. Silva, "Exploring retrospective annotation in long-videos for emotion recognition," *IEEE Trans. on Affective Computing*, vol. 15, no. 3, pp. 1–12, 2024

SJR: Q1; IF: 9.6

Obj 3. Emotion Annotation for Naturalistic Settings

10. P. Bota, J. Brito, A. Fred, P. Cesar, and H. Silva, "A real-world dataset of group emotion experiences based on physiological data," *Scientific Data*, vol. 11, no. 1, pp. 1–17, 2024

SJR: Q1; IF: 5.8

Obj 4. Real-world Affective Computing Dataset

11. L. Aly, L. Godinho, P. Bota, G. Bernardes, and H. P. da Silva, "Acting emotions: A comprehensive dataset of elicited emotions," *Scientific Data*, 2024

SJR: Q1; IF: 5.8

Obj 4. Real-world Affective Computing Dataset

1.3.2 International Conferences

1. P. Bota, C. Wang, A. L. Fred, and H. Silva, "A wearable system for electrodermal activity data acquisition in collective experience assessment," in *Proc. of the Int'l Conf. on Enterprise Information Systems*, pp. 606–613, 2020

CORE: C

Obj 2. Group-based Physiological Data Collection

2. G. F. D. Salvador, P. Bota, V. Vinayagamoorthy, H. Plácido da Silva, and A. Fred, "Smartphone-based content annotation for ground truth collection in affective computing," in *Proc. of the Int'l Conf. on Interactive Media Experiences*, p. 199–204, ACM, 2021

Obj 3. Emotion Annotation for Naturalistic Settings

3. P. Bota, A. Fred, J. Valente, and H. Silva, "Automatic classification of physiological signals modalities," in *Int'l Meeting of the Portuguese Society of Physiology*, 2019

Obj 4. Real-world Affective Computing Dataset

4. L. Aly, P. Bota, L. Godinho, G. Bernardes, and H. Silva, "Acting emotions: physiological correlates of emotional valence and arousal dynamics in theatre," in *ACM Int'l Conf. on Interactive Media Experiences*, pp. 381–386, 2022

Obj 4. Real-world Affective Computing Dataset

1.3.3 Student Guidance and Supervision

I have co-supervised students completing training programs over 4 MSc, 5 BSc and 6 intern students, ranging from European Qualifications Framework (EQF) Level 5 (BSc) to EQF Level 6 (MSc), from diverse backgrounds from biomedical to electrical engineering and computer science, with works producing technical and scientific achievements of international relevance, including journals and international conferences.

Due to the institutional requirements for student supervision, most cases have been performed in collaboration without being a formal co-advisor, these are identified by the *.

Master Students

1. S. Morgado, "AI-powered emotion well-being for everyone," MSc thesis, Instituto Superior Técnico da Universidade de Lisboa, 2024.
Advisor: Sangra Gama (IST-UL/INESC-ID) and Hugo P. da Silva (IST-UL/IT)
2. *J. Alves, "Facial emotion recognition for mental well-being assessment in the workplace," MSc thesis, Instituto Superior Técnico da Universidade de Lisboa, 2022.
Advisor: Hugo P. da Silva (IST-UL/IT) and Ana Fred (IST-UL/IT)
3. *G. Salvador, "Real world group emotional analytics using eletrodermal activity signals," Master's thesis, Instituto Superior Técnico, Universidade de Lisboa, 2021
4. *C. Lima, "Psychophysiological effects of guided imagery based intervention on the academic development of children," MSc thesis, Faculdade de Ciências da Universidade de Lisboa, 2024.
Advisor: Hugo P. da Silva (IST-UL/IT), Nuno Matela (FCUL) and Brigida Ferreira (FCUL)

Bachelor Students

1. T. Talento, "Movie emotional content analysis," 2023. BSc final project, Instituto Superior Técnico - Universidade de Lisboa
Co-Advisor: Hugo Silva (IST-UL/IT)
2. I. Salema, "Emotion analysis using facial expression recognition," 2023. BSc final project, Instituto Superior Técnico - Universidade de Lisboa
Co-Advisor: Hugo Silva (IST-UL/IT)
3. M. Supelnic, "Benchmarking of sensor configurations and measurement sites for out-of-the-lab photoplethysmography," 2024. BSc final project, Instituto Superior Técnico - Universidade de Lisboa
Co-Advisor: Afonso Ferreira (INESC), Hugo Silva (IST-UL/IT)
4. F. Silva and Z. Xu, 2024. BSc final project, Instituto Superior Técnico - Universidade de Lisboa
Co-Advisor: Hugo Silva (IST-UL/IT)
5. S. Silvestre, "Learning enhancement using affective computing," 2024. BSc final project, Instituto Superior Técnico - Universidade de Lisboa
Co-Advisor: Hugo Silva (IST-UL/IT), Cátia Costa (HLuz), José Moreira (HLuz)

Interns

1. J. Brito, "A proof-of-concept neurosecurity study based on peripheral physiological signals," 2022 – 2023. Technical Report, Instituto Telecomunicações
2. A. Gonçalves, "Integration of ScientISST core in EmotiphAI," 2022. Instituto Telecomunicações
3. P. Correia, "Real-world data collection," 2021 – 2022. Instituto Telecomunicações
4. C. Bento, "Data collection using the FMCI device," 2019 - 2020. Instituto Telecomunicações
5. D. Venancio, "Kubernetes for group data collection - EmotiphAI," 2023. Instituto Telecomunicações
6. F. A. Santos, "Firmware and hardware developer - ScientISST," 2023 - 2024. Instituto Telecomunicações

1.3.4 Dissemination Activities

In addition to academic activities, I was involved in outreach activities such as workshops, talks, and scientific fairs, underscoring my dedication to spreading biomedical engineering, enhancing scientific understanding, and engaging with the community.

Invited Speaker

1. P. Bota and d. S. H. P., "Emotion assessment in the wild," in *Modern Technologies Enabling Innovative Methods for Maritime Monitoring and Strengthening Resilience in Maritime Critical Infrastructures*, (Lisbon, Portugal), Jan 2024
2. P. Bota and A. Ferreira, "What do your biosignals say about you," in *INBIO (Introdução aos Biosinais)*, (Lisbon, Portugal), Jul 2023
3. P. Bota, M. Abreu, and H. P. da Silva, "Introduction to machine learning and applications," in *Exercise Prescription and Health Promotion*, (Instituto Politécnico de Leiria, Leiria, Portugal), Feb 2020

Workshops

1. P. Bota and J. Brito, "EmotiphAI: How to detect emotions in a group of people," in *Maker Faire*, (Lisbon, Portugal), Feb 2023
2. P. Bota, H. P. da Silva, and M. Abreu, "Introduction to biosignal acquisition," in *1st Int'l Meeting of the Portuguese Society of Physiology*, (Lisbon, Portugal), Oct 2019

3. P. Bota and M. Abreu, "Signal processing and machine learning," in *summer course of CEiiA (Centre of Engineering and Product Development)*, (Instituto Superior Técnico, Lisbon, Portugal), Jul 2020
4. P. Bota and M. Abreu, "Introduction to machine learning and applications," in *Clynx*, (online), Jul 2020
5. P. Bota and M. Abreu, "Physiological signal classification," in *Congress of the Brazilian Society of Physiology*, (online), Sep 2020

Scientific Fairs

1. P. Bota, A. S. Carmo, M. Abreu, and S. Monteiro, "EmotiphAI and IT group projects showcase." Noite Europeia dos Investigadores, Sep 2023
2. P. Bota and A. M., "EmotiphAI and IT group projects showcase." Dia Internacional das Raparigas nas Tecnologias de Informação e Comunicação, Apr 2023
3. P. Bota, J. Brito, R. Silva, and V. Garção, "Biosignals acquisition and visualisation." FIC.A (International Festival of Science), Oct 2022
4. P. Bota, S. Monteiro, L. Pereira, R. Silva, and R. Maciel, "EmotiphAI and it group projects showcase." 22nd anniversary of Campus Taguspark, Nov 2022
5. P. Bota and M. Abreu, "Dissemination of FMCI xinhuanet signal acquisition devices." Web Summit, Lisbon, Portugal, Nov 2019.
In collaboration with Xinhuanet FMCI

Showcases

14 Jun 2022 Online theatre as a research instrument

Descrip.: Experiment to find a method of collecting feedback from viewers on an online event. I was involved in developing the infrastructure for the audience's physiological data collection.

Team: Taavet Jansen (UTallinn), Aleksander Väljamäe, Joana Brito (IT), Hugo Silva (IST-UL/IT), and Ana Fred (IST-UL/IT).

2020 Opera Experience

Descrip: Experiment to analyze an audience response to a live Opera show at ESMAE. I was involved in developing the infrastructure for the audience's physiological data collection and the design of the experiment.

Team: Ana Rosado (ESMAE), António Salgado (ESMAE), Hugo Silva (IST-UL/IT), and Ana Fred (IST-UL/IT).

14-19 Oct 2019 DataWe

Descrip.: A neurocinematic system for studying collective decision-making shown at the European Commission Joint Research Centre Ispra Sector, Italy. I was involved in developing the infrastructure for the audience's physiological data collection and communicating with the media feedback.

Team: Taavet Jansen (UTallinn), Aleksander Väljamäe (Univ. of Tartu), Anu Almik, Hugo Silva (IST-UL/IT), and Ana Fred (IST-UL/IT).

1.4 Outline

This thesis is structured into eight chapters, each addressing a different component in an end-to-end affective computing pipeline as depicted in Figure 1.1.

EmotiphAI

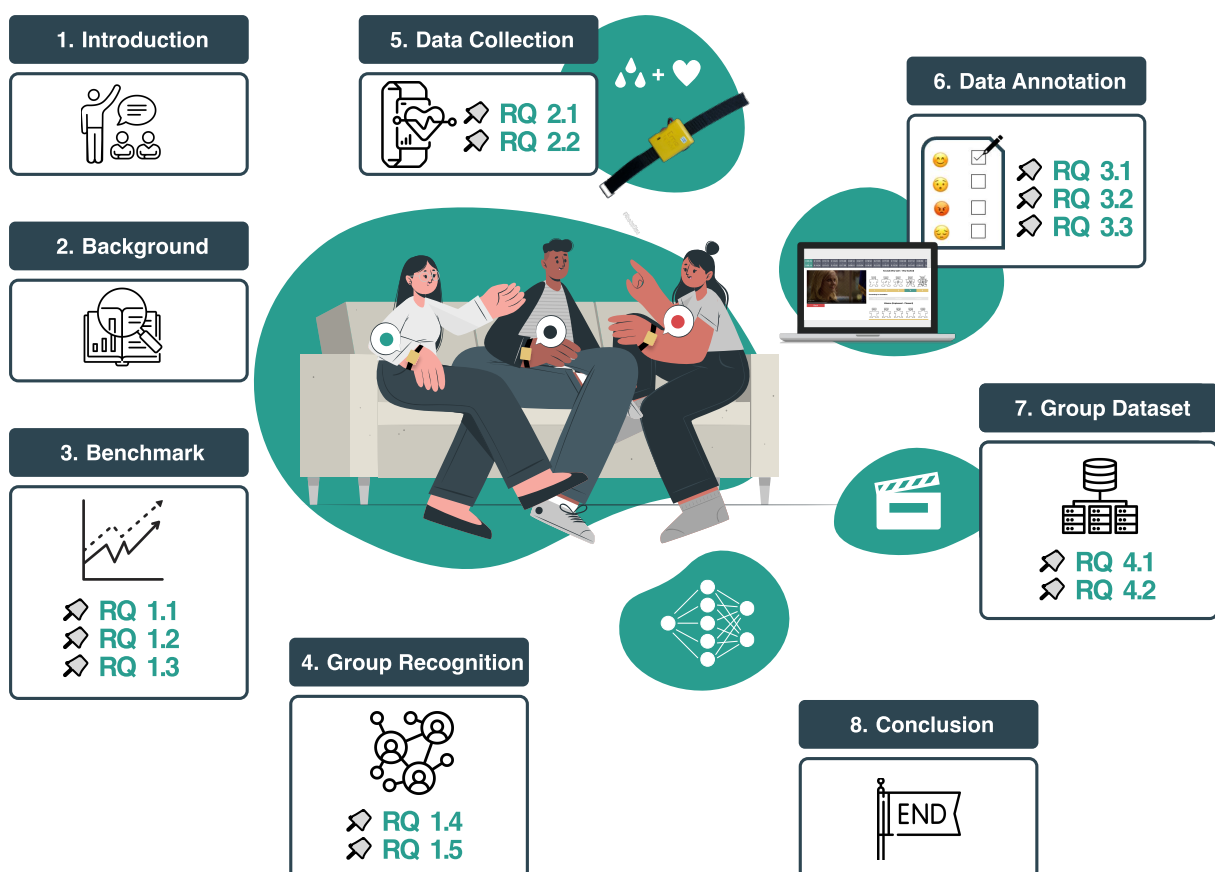


Figure 1.1: Thesis content overview.

Towards the end goal of group emotion recognition, Chapter 2, conducts a thorough literature review, exploring the underpinnings of emotion theory (Section 2.1), the methods for measuring emotion (Sec-

tion 2.2), existing datasets (Section 2.3), and the influence of social interaction on modulating emotion (Section 2.4).

The thesis is then divided to address the four main objectives presented in Section 1.1, and their respective RQs.

Chapter 3 addresses *Obj 1. on the Development of Affective Computing Algorithms*, namely RQ 1.1 to 1.3 by reviewing the state-of-the-art in emotion recognition and analysing the prevailing physiological sensors, algorithms and datasets. This chapter also evaluates the performance of various classifiers and feature sets, benchmarking the expected performance against public datasets.

Chapter 4 further addresses *Obj 1. on the Development of Affective Computing Algorithms*, namely RQ 1.4 and 1.5 by demonstrating on a public dataset, how group-related information can be leveraged to enhance emotion recognition systems.

Chapter 5, addresses *Obj 2. Group-based Physiological Data Collection*, namely RQ 2.1 and 2.2 by introducing the EmotiphAI platform, consisting of the EmotiphAI Wearable, Data Collector and User Interface designed for pervasive individual and group physiological data acquisition. This chapter also presents the results of a study on the impact of the data collection sampling rate and network technology on the quality of the multi-device data collection.

Considering that the development of emotion recognition systems requires not only physiological data but also emotion annotations, Chapter 6 addresses *Obj 3. Emotion Annotation for Naturalistic Settings*, namely RQ 3.1 to 3.3 by presenting the EmotiphAI Annotator. The latter consists of an emotion annotation tool designed for the annotation of physiological data within naturalistic environments, i.e. closer to real-life experiences. This chapter also presents the results of a study on the usability and reliability of the EmotiphAI Annotator.

Building upon the data acquisition and annotation tool developed in the previous chapters, Chapter 7 addresses *Obj 4. Real-world Affective Computing Dataset*, namely RQ 4.1 and 4.2 by presenting G-REx, a dataset for group emotion recognition in a naturalistic setting. Consolidating, how the EmotiphAI platform can be used to collect and annotate large amounts of physiological data in a naturalistic setting.

Finally, Chapter 8 provides a summary of key research findings, explores insights gathered from the field, and identifies future challenges and possible directions that lay ahead in further developments in group emotion recognition through physiological signals.

Chapter 2

Background

Grasping the essence of emotion and its measurement techniques is crucial for understanding and developing emotion recognition systems. Towards the end-goal of *Obj 1.*, focused on the *Development of Affective Computing Algorithms*, this chapter starts by outlining the principal theories of emotion, then breaks down emotion into its fundamental components, and explains how various emotions can be quantified, with a special focus on physiological signals. The discussion then details the existing datasets for emotion analysis through physiological signals available in the literature. The chapter concludes by examining the interplay between group emotions and individual affect, highlighting how the group setting and its emotional dynamics can shape and influence emotions at the individual level.

Contents

2.1	Emotion Theories	14
2.1.1	Basic Emotion Theory	14
2.1.2	Appraisal Theories of Emotion	15
2.1.3	Psychological Construction Theory of Emotion	16
2.1.4	Multicomponent Theory of Emotion	17
2.1.5	Discussion	19
2.2	Measuring Emotions	20
2.2.1	Subjective Experience	20
2.2.2	Central Nervous System	22
2.2.3	Behaviour	24
2.2.4	Autonomic Nervous System	26
2.2.5	Discussion	32
2.3	Datasets	33
2.4	Group Emotions	35
2.4.1	Groups	36

2.4.2	Collective Emotions	36
2.4.3	Emotion Transmission in Groups	38
2.4.4	Discussion	40

2.1 Emotion Theories

Since Ancient Greece [50] up to today, philosophers have wondered and struggled with the concept of emotions. Fehr and Russell famously said [51]:

Everyone knows what an emotion is until asked to give a definition. Then, it seems, no one knows.

Over the years, multiple theories have evolved to explain the underlying structure, origin and purpose of emotions.

2.1.1 Basic Emotion Theory

The classic view on emotion builds upon Darwin's evolutionary perspective [52]. Darwin posed that emotions evolved as adaptive mechanisms with crucial roles in survival, such as "avoiding predators, finding food, or caring for offspring" [53]. In this view, emotions regulate the interactions with the environment (e.g. the feeling of disgust might prevent one from eating something rotten, and the feeling of fear might make one widen the eyes, increasing the field of vision and preparing the body to "fight-or-flight") [1], and/or serve as effective means of communication towards other peers or threads [54] (e.g. a blush might show honesty and a tear sadness).

The basic emotion theories are in line with the classical view, with emotions seen as evolutionary byproducts, ingrained within our genetic makeup. These theories propose that emotions have arisen consistently across diverse cultures and have endured through time, being a universal phenomenon. Given their innate and ubiquitous nature, each emotion category is associated with a distinctive "fingerprint", a pattern of brain activation that is consistent and replicable across different populations [55].

Perhaps an annoying coworker triggers your "anger neurons", so your blood rises, you scowl, yell, and feel your heart race; you freeze and feel a flash of dread [55].

The fingerprints of emotions are observed for a set of basic (Ekman [56]), primary (Plutchik [57]), discrete (Izard [58]), and affect programs (Tomkins [59]) emotions, which are the building blocks for more complex emotional states. Across theorists, the number of basic emotions and selection requirements for this nomenclature varies, however common agreements denote that basic emotions should be universally expressed, accompanied by a unique subjective experience, and they fulfil adaptive roles that have been shaped by evolution [60].

Table 2.1 shows some of the predominant theorists that adopt the basic theory of emotion and their proposed set of basic emotions. Although the nomenclature and number of basic emotions vary, overall, six basic emotions are common across the literature, namely: happiness, sadness, anger, disgust, fear, and surprise.

Table 2.1: List of predominant basic emotions in the literature. Table adapted from [60].

Reference	Happiness (Joy, Enjoyment, Play)	Sadness (grief)	Anger (Rage, Hatred)	Disgust	Fear (Anxiety)	Surprise/ Wonder	Other
Plutchnik [57]	✓	✓	✓	✓	✓	✓	Acceptance
Oatley and Johnson-Laird [61]	✓	✓	✓	✓	✓	✓	
Ekman and Cordaro [62]	✓	✓	✓	✓	✓	✓	Contempt
Izard [58]	✓	✓	✓	✓	✓	✓	Interest
Levenson [63]	✓	✓	✓	✓	✓	✓	Interest, Relief, Love
Panksepp and Watt [64]	✓	✓	✓	✓	✓	✓	Seeking, Lust, Care
Tomkins:2006 [59]	✓	✓	✓	✓	✓	✓	Interest, Contempt,
Descartes [65]	✓	✓	✓	✓	✓	✓	Distress, Shame
Arnold [66]	✓	✓	✓	✓	✓	✓	Desire, Love
							Desire, Despair, Hope, Love, Courage, Dejection

Much of the knowledge of basic emotion theory has been derived from studies on emotional facial expressions, most notably those conducted by Ekman [62]. Ekman identified a set of microexpressions associated with a range of categorical emotions: the Facial Action Coding System (FACS) [67]. However, despite its contributions, basic emotion theory has been met with criticism and scepticism, including the lack of consistency for distinct neural and peripheral physiological responses across basic emotions [60, 68], and the existence of contradictory tests on the universality of emotional facial expressions [1].

Nevertheless, the classic view of emotion is still very much present in today's society and culture, for example in TV shows like "Lie to Me" which base their plot on the idea that one inner thoughts can be read by facial and corporal expressions, or in companies like Affectiva which analyse body reactions to measure metrics such as stress and engagement to improve advertisement/contribute to road safety [55].

2.1.2 Appraisal Theories of Emotion

Appraisal theorists argue that emotions are not automatic reactions to stimuli, rather they are shaped by the individual's subjective interpretation and unconscious evaluation over a set of appraisal components. As defined by Lazarus [69], appraisal is the cognitive process involved in emotion elicitation. The cognition component evaluates: 1) if the stimulus leads to an emotional state; 2) which emotion should be elicited; and 3) the emotion's intensity [70]. The appraisal components differ from theorist to theorist, however, common agreement is established in: "*novelty, intrinsic pleasantness (or unpleasantness), predictability, goal-relevance (i.e. their significance for the individual's goals), coping potential (i.e. the individual's ability to cope with the consequences of the event), and normative significance (i.e. the compatibility with personal or social norms and values)*" [1].

Within the appraisal theories, two sub-families can be identified: causal/classical appraisal; and the

constitutive/constructive appraisal theories [1]. Similarly to basic emotion theory, causal appraisal is an Essentialism theory, while constitutive is not. Essentialism [71], is the belief that emotion instances that belong to the same category (e.g. happiness) have a shared pattern that describes them and unique underlying mechanisms that originate them. In the causal theory, appraisal variables have unique internal mechanisms that can be combined to form a distinctive appraisal pattern defining the emotion category [71]. On the other hand, in constructive appraisal theories, appraisals are not causal antecedents of emotions but constituents, i.e. the emotion components are dependent on the cognitive, perceptual and situational content, creating variability across and within emotion categories [3].

Basic emotions too can be associated with evolutionary appraisal variables: happiness with subgoals being achieved; anger with frustration in active plans; sadness with failure of a major plan or loss of active goal; fear with self-preservation goal threatened or goal conflict; and disgust with a gustatory goal violated [53]. However, the two theories are separated by cognition, social, linguistic and cultural factors taking part in the definition of the emotional experience in the case of appraisal theories [3]. As emotions diverge towards subjective cognitive processes, they demonstrate greater cultural and species-specific variations, departing from the basic emotion theory and common universal patterns [72].

2.1.3 Psychological Construction Theory of Emotion

Upon the difficulties in identifying unique patterns for emotion categories, alternative views to the Essentialism view (where each emotion category has a set of attributes necessary to its identity and function) [71] have emerged. An example is the theory of constructed emotion, where rather than emotions being pre-programmed universal reactions to stimuli, they are constructed by the individuals as they take their interior and exterior sensations. The individual past and the event's cultural meaning are used to conceptualize the emotion, leading to a unique pattern to that stimulus and individual. An example is given by Lisa Feldman Barrett [55]:

"When you are angry, you might scowl, frown mildly or severely, shout, laugh, or even stand in eerie calmness, depending on what works best in the situation. Your heart rate likewise might increase, decrease, or stay the same, whatever is necessary to support the action you are performing."

The constructed view maintains the idea of emotion categories, with emotion instances belonging to emotional categories (e.g. Anger). However, variability is the norm as one emotion category can have many manifestations and forms according to what better response fits the situation. The context where the stimulus happens, the individual's past experiences and culture will cause the variation observed in the emotion category entities.

Emotion constructivist models are characterized by two components. The first consists of the Core

Affect: the individual's neurophysiological state that changes as it receives internal and external sensory input. The second consists of a categorisation/perception step with the brain constantly interpreting bodily sensations based on past experiences and the individual's culture to predict how to better respond to the stimulus. The Core Affect is available to consciousness and is categorised using the valence-arousal dimensions [73]. The valence-arousal sensation is then categorized by the brain using linguistic concepts learned through cultural exchanges. Valence is described as a subjective experience of pleasure or displeasure, while arousal is a state of activation or deactivation (sense of energy mobilization). Unlike emotion, valence is considered universal across the literature, with the consensus that since birth we can differentiate the feelings of pleasure or displeasure/comfort or pain [74]. Russel [75], describes valence and arousal as *"states of the central nervous system, both have peripheral physiological correlates and both are subjectively experienced as mental events"*. Alternative dimensions such as approach and withdrawal [76], tension and energy [77], negative activation and positive activation [78], have been shown to describe similar processes as arousal and valence [79].

The empirical support for the valence-arousal dimensions is large: 1) Valence and arousal have shown to be primitive and universal [80]; 2) Self-reports across different time frames, languages, facial expressions and discrete works were able to reproduce the 2D structure of emotion [68]; and 3) A direct correlation has been observed between the two factors and peripheral physiological responses, e.g. an increment of skin conductance and HR has been correlated with an increase of arousal [81]. Facial muscles, including the corrugator muscle associated with negative emotions and the zygomatic muscle linked to positive emotions, have been found to correlate with changes in emotional valence; specifically, an increase in negative valence is associated with the corrugator muscle, while an increase in positive valence correlates with the zygomatic muscle [81]. Similarly, vocal characteristics have been shown to be indicative of arousal levels [5].

Overall, the theory of constructed emotion offers a nuanced approach to measuring emotion. It incorporates emotional self-reports, a multi-modal measuring system covering autonomic measures, facial expressions, the individual experience and perception, as well as the stimulus context [71]. This multi-faced view of emotion elevates the task of automatic emotion recognition to higher levels of complexity.

2.1.4 Multicomponent Theory of Emotion

In the work by Paul and Anne Kleinginna [82], the authors analyzed nearly 100 different emotion definitions and divided them into 11 separate categories. Despite the diversity in the theories of emotion, common diagnostic features can be identified, formulating the componential theories of emotion [83]. An illustrative example of the components involved in an emotional experience was given by Chat-GPT¹:

"Suppose you're at a family gathering, and an estranged relative whom you haven't seen

¹<https://openai.com/chatgpt>; Accessed on 20/02/2024

for years makes an unexpected and unwelcome appearance. This relative has a history of making condescending remarks toward you. Within moments of their arrival, they make a snide comment about your appearance. Let's dissect the emotional episode of indignation or resentment that this might trigger:

Cognitive Appraisal: You interpret their comment as another in a long line of unjust and derogatory remarks aimed at belittling you, deepening the rift between you two.

Physiological Changes: You might feel a rush of heat to your face, a tightening of your throat, and a quickening of your pulse. Your palms might become sweaty, and your stomach might churn in response.

Distinctive Expressions: Your facial expression becomes stern, with eyes narrowing and a frown forming. Your posture stiffens, and your voice takes on an icy, restrained tone as you respond.

Subjective Experience: Internally, you feel a mix of hurt, anger, and determination. There's a strong desire to either confront the relative or to disengage and distance yourself from them to avoid further conflict.

Mental Processes and Behavioral Dispositions: Your entire focus shifts from enjoying the gathering to this interaction. Instead of mingling and catching up with other loved ones, your thoughts revolve around past incidents with this relative and ways to cope with the current situation. You might consider leaving early, seeking support from another family member, or directly addressing the insult."

The example illustrates some of the main characteristics of an emotional episode: how emotions are powerful phenomena, being the driving force behind motivation, modulating your thoughts, cognition and ultimately, decision-making; how the initial cognitive appraisal of the situation is dependent on one's experiences, background, and culture, i.e. upon an equal situation, another individual might experience a completely different emotion and act differently; and lastly, the example shows some diagnostic features of emotion that let you conclude you are experiencing an emotion.

A consensual definition in the emotion literature establishes emotion as an event-focused process from a relevant stimulus event (internal or external) to which it follows a synchronized multi-component response mobilizing resources towards an action tendency that responds to the stimulus [1, 4]. The multi-component view of emotion includes cognitive appraisal where the subject evaluates the stimulus; physiological processes such as e.g. electrodermal and cardiovascular responses that prepare the individual's body to better respond to the stimulus; body expressions such as facial, voice and body changes that act as a communicative tool about the individual's intentions to respond to the stimulus; and lastly, the subjective emotional experience/feeling [1, 84]. The selection of the diagnostic features,

order and magnitude are some of the building blocks that distinguish different theories of emotion.

Since emotions require an intense and synchronized response mobilization across diverse systems, they must be short-lasting [83]. This definition separates emotion from a baseline emotional state, the individual mood: a prolonged subjective feeling of low intensity, which does not result from any specific event and without elevated resource mobilization [83].

2.1.5 Discussion

The main emotion theories described previously can be related by their range of Essentialism [71] (see Section 2.1.2). Basic emotion and causal appraisal theories (Darker green in Figure 2.1) are built upon Essentialism. However, while for basic emotion theories emotions derive from neural and physiological specific circuits, in causal appraisal theories, emotions are formed by specific patterns of appraisal variables [1]. According to essentialist theories, emotions have unique fingerprints, thus, they can be measured and identified by facial expressions, body responses, the autonomic nervous system or the central nervous system [71]. The opposite view is taken by the physiological construction theories (lighter green in Figure 2.1), to which no unique pattern is observed in emotion categories since emotions are created in-loco based on the environment and how the individual makes sense of how to better react to the stimulus. This overview of the literature illustrates the varied perspectives on emotion, highlighting both the diversity and how emotion theories intersect.

Characteristics from diverse theories of emotion are borrowed throughout the remainder of this work, specifically in Chapters 3, 6 and 7. A search for distinct "fingerprints" for the development of automatic emotion recognition systems necessitates alignment with the Essentialist perspective of emotion. The identification of these fingerprints relies on the multi-component view of emotion, involving measurements of the Autonomic Nervous System (ANS) peripheral signals and the individuals' self-reports on

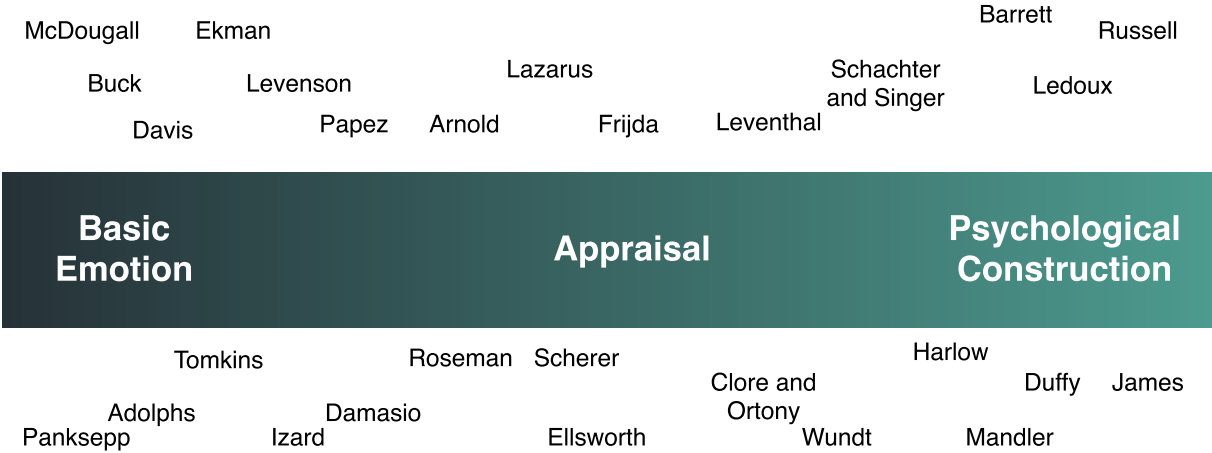


Figure 2.1: Emotion theories are grouped into three main families: Basic Emotion, Appraisal and Psychological construct theories. Figure adapted from [1].

their subjective experience. Moreover, the latter is described on the valence and arousal dimensions based on the constructionist theories. Lastly, in line with the theory of constructed emotion, the environment is captured by incorporating the group context in Chapter 4, with the goal of obtaining a more holistic representation of emotion.

2.2 Measuring Emotions

A consensual view of emotion posits that the process begins with sensory inputs, which are monitored by brain regions such as the amygdala. This is followed by a rapid and unconscious appraisal to determine whether these inputs pose a threat to the organism's survival or overall well-being. Subsequently, the CNS may interrupt ongoing activities and, through the ANS, reallocate physiological resources to address the perceived threat or challenge [85].

The componential view of emotion suggests that the emotion components (Table 2.2) can be analyzed to capture and measure emotions, enabling the comparison of the various emotion theoretical approaches.

Table 2.2: Emotion components, their measurement methods and reported dimensions. Table adapted from [5, 83].

Response System	Measure	Sensitivity	Function
Subjective Experience	Self-Report	Valence and Arousal, Basic Emotions	Monitoring of Internal State and Environment Interaction
Peripheral Physiology (ANS)	ANS	Valence and Arousal	Regulation
Central Physiology (CNS)	fMRI, PET, EEG	Approach and Avoidance	Evaluation, Preparation and Direction of Action
Behaviour	Vocal characteristics, Facial behaviour (observer ratings), Facial behaviour (EMG), Whole body behaviour	Arousal, Valence, Emotion Specificity	Communication of Reaction, Behavioural Intention

2.2.1 Subjective Experience

The analysis of subjective experience namely through user questionnaires is the most predominant measure in the emotion literature. Moreover, it is the gold standard in medical or psychotherapy practices as well as the ground truth used to analyse peripheral, central or behavioural responses.

The emotional assessment can be performed either by the subjects themselves (self-assessment) or by an external individual (implicit assessment) through the observation of external body behaviour. Both present challenges. Starting with self-assessment, the method is dependent on the subject inferring their

own emotions and transferring them to the possibilities presented in the questionnaires. Self-reports can be biased as the individual has to approximate their emotion conceptualization (e.g. emotion of excitement) to the self-report scale (which can only have happy/sad as an option) [71]. Additionally, self-assessment can be felt by the subject as an intrusive process, evoking a defence mechanism with the subject providing an unreliable report of their emotions both unconsciously or consciously to preserve their privacy [86].

On the other hand, in external annotation, independent subjects assess the subject's affective state through the analysis of their externally observable behaviour [87]. This means that the annotator is dependent on the subject's ability or willingness to externalize their emotions. Such factors can be correlated with the subject's personality, environment and/or culture.

A debated topic is how can the subject describe its subjective experience. Research in this area is supported by the different theories of emotion, presenting questionnaires to the subject of discrete emotions (following the basic theories [62]), appraisal variables (following the appraisal theories [83]), or core dimensions (following the core affect from the physiological constructivists/dimensional theories [75]).

The literature has validated scales for assessing emotional responses, common to numerous studies and datasets. A predominant scale observed in the literature datasets [87, 88, 89] for emotion recognition is the SAM scale [90]. SAM is a graphical scale designed to facilitate the cross-cultural measurement of emotional response by providing the emotional dimensions of valence, arousal and dominance through manikins (Figure 2.2). The SAM is usually filled out after the emotion elicitation, i.e. retrospectively.

An alternative to SAM identified in the literature [87, 88] is PANAS [91], consisting of two 10-item mood scales, one measuring positive affect and the other measuring negative affect. The positive affect axis reflects enthusiasm, activation and alertness, while the negative axis reflects distress and unpleasant engagement.

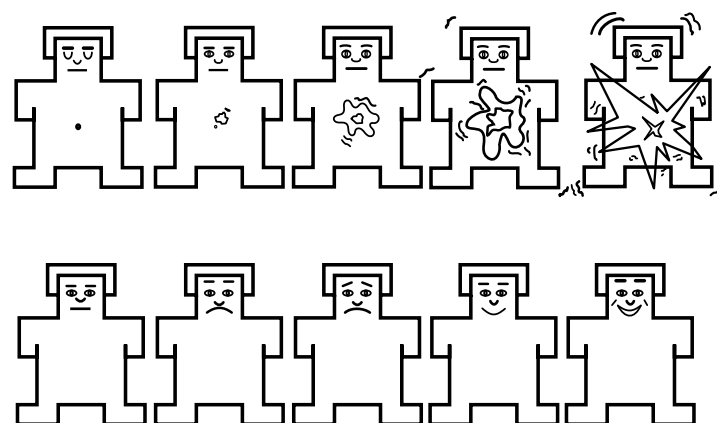


Figure 2.2: SAM scale for the arousal (first row) and valence (second row) in a 1 to 5 scale [90].

Based on these scales, diverse platforms have been tested for emotion annotation which will be further analyzed in Chapter 6.

2.2.2 Central Nervous System

The brain is involved in generating and modulating emotions. Information regarding which brain regions are activated upon emotional events can be used to identify the existence of distinctive neural patterns in emotion categories. The measurement of specific brain region activations is typically performed using physiological sensors (e.g. EEG) or imaging techniques (e.g. fMRI, or PET).

The EEG measures the electric field created by currents that flow when neurons are activated [9]. To analyze the activated brain regions, the EEG signal is often decomposed in five frequency bands: Delta (< 4 Hz), Theta (4 – 7 Hz), Alpha (8 – 13 Hz), Beta (14 – 30 Hz), and Gamma waves (31 – 100 Hz) [92]. Research has indicated that higher frequencies, particularly in the Gamma and Beta ranges, carry useful information for emotion recognition [92]. For instance, Li and Lu analyzed Gamma waves to distinguish between happiness and sadness, achieving an accuracy exceeding 90% [93].

The EEG has high temporal resolution, but low spatial resolution. Due to the limitations in spatial resolution, a common approach is to focus on left versus right-side or anterior versus posterior activation of the brain. Following this approach, the literature using the EEG has recorded patterns of emotion asymmetry. For example, initial studies as the one by Davidson *et al.* [94] linked left-side activation in the anterior temporal region with positive-valence stimuli, and the right-side activation in the frontal and anterior temporal region with negative stimuli [5]. Later studies by Davidson and Sutton [95], identified prefrontal asymmetry to be linked with approach-avoidance dimensions, with the left-side activation linked to approach and the right-side activation linked to avoidance.

In contrast to the EEG, neuroimaging techniques (e.g., fMRI or PET) have a high spatial resolution but low temporal resolution. Thus, they can be key in identifying specific brain regions related to different emotion categories. The fMRI measures the level of oxygenation in the brain through oxygenated Hemoglobin (Hb), which is a strong magnetic material. When a part of the brain is more active, it requires more oxygen, so blood flows to that region increasing the volume of oxygenated Hb. The PET technique, on the other hand, measures the radiation emitted by a radioactive substance, which accumulates in areas with higher activity. Both techniques can detect activated brain regions through blood flow [96].

The experimental results using imaging techniques support the EEG results of neural asymmetry related to dimensional theories of emotion. Namely, in the work by Daly *et al.* [97], the authors reported prefrontal asymmetry related to approach – withdrawal behaviour while listening to music. However, the results are not clear as demonstrated by Murphy *et al.* [96], where the approach dimension was related to left-sided activity, but negative/withdrawal was found to be symmetrical.

More contemporary research has taken advantage of imaging techniques for higher spatial resolution and focused on identifying specific brain regions (Table 2.3). These areas are visually depicted in Figure 2.3, where the author identified the brain activation areas that most contributed to the classification of six basic emotions [98].

Table 2.3: Summary of brain regions identified across emotion theories [1, 99, 100, 101, 102, 103].

Theory	Emotion	Brain Region
Basic	Fear	Amygdala
	Disgust	Insula, Ventral Prefrontal Cortex and Amygdala
	Sadness	Medial Prefrontal Cortex and Anterior Cingulate Cortex
	Anger	Orbitofrontal Cortex
	Happiness	Rostral Anterior Cingulate Cortex, Ventromedial and Prefrontal Cortex
Appraisal	Novelty	Hippocampus
	Relevance	Amygdala
	Goal Incongruence	Anterior Cingulate Cortex
	Agency	Temporal Parietal Junction
	Norm Compatibility	Dorsolateral Prefrontal Cortex
Constructivist	Arousal	Amygdala
	Valence	Orbitofrontal Cortex

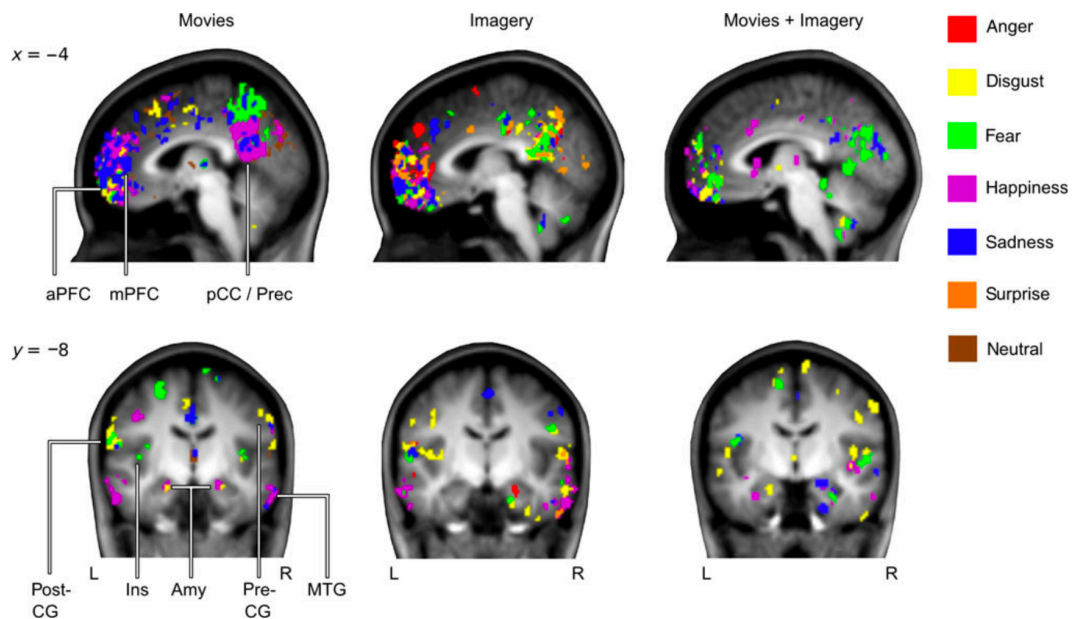


Figure 2.3: Brain regions correlated to 6 basic emotions in the work by Heini *et al.* [98]: aPFC – anterior prefrontal cortex; Amy – amygdala; Ins – insula; LOC – lateral occipital cortex; mPFC – medial prefrontal cortex; MTG – middle temporal gyrus; PCC – posterior cingulate cortex; post-CG – postcentral gyrus; pre-CG – precentral gyrus; Prec – precuneus.

Overall, the literature [1, 99, 103] on CNS correlates to emotion, has shown that emotions seem to be associated with the activation of cortical and sub-cortical areas, with no contribution in a one-to-one emotion, rather, each region is activated for separate emotions. Moreover, the evidence of one-to-one

correspondence between emotions and brain region as the basic emotion theory originally defended is insufficient. In turn, there is a consensus that more complex networks are necessary, such as functional networks where one (emotion)-to many (neural regions) is identified [103]. For a more comprehensive understanding of the underpinnings of neural regions related to emotion, the reader is referred to [54, 99, 104].

2.2.3 Behaviour

An additional component of emotion involves the body's behaviour, either through facial expressions, whole-body actions, gestures, postures, or changes in vocal expression [5]. Darwin [52] advocated that these body-related changes were functional tools preparing our organism to better respond to the emotion-induced stimulus [1], and/or are a form of communication to others [54]. All major theories of emotion include behaviour clues, either to express or prepare action tendencies [1]. In our daily lives, we subconsciously perform emotion-induced behaviour changes, and often resort to these changes as clues to perceive the emotions of others.

The most explored component of emotional body behaviour is facial expressions. The creation of prototype images of facial expressions depicting basic emotions was first introduced by [105], and later disseminated by Ekman in its cross-cultural studies of facial expressions [67]. Ekman reported the existence of over 42 muscles in the face, which can be combined to create over 21 macro-expressions (with a duration of around 0.5 to 4 seconds). Simultaneously to macro-expressions, the literature reports the existence of micro-expressions, however, smaller in duration (around 0.03 seconds) [84].

Through the analysis of micro-expressions, Ekman [67] developed the Facial Action Coding System specifying facial muscle movements, denoted as action units involved in human facial expression that allow the identification of six basic emotions (anger, fear, disgust, happiness, sadness and surprise). These emotions were considered universal since they could be recognized across different cultures [5]. An example with seven common action units from the upper part of the face is shown in Figure 2.4, illustrating how action units act as building blocks to create diverse facial expressions.

An alternative to using the human eye for facial emotion recognition is the use of computer-based measuring systems, such as the EMG, which enables higher precision in the detection of minute movements.

The EMG is a technique that monitors muscle activity by measuring the electrical signals generated during muscle contraction and relaxation. When the brain or spinal cord wants to initiate a muscle movement, it sends out an electrical signal that travels through the nervous system, activating motor neurons and generating an action potential in the muscle fibres it innervates. The action potential triggers the release of neurotransmitters, which lead to muscle contraction. When the stimulation of the motor neuron stops, the muscle relaxes and returns to its resting state.
















<i>NEUTRAL</i>	AU 1	AU 2	AU 4	AU 5
				
Eyes, brow, and cheek are relaxed.	Inner portion of the brows is raised.	Outer portion of the brows is raised.	Brows lowered and drawn together	Upper eyelids are raised.
AU 6	AU 7	AU 1+2	AU 1+4	AU 4+5
				
Cheeks are raised.	Lower eyelids are raised.	Inner and outer portions of the brows are raised.	Medial portion of the brows is raised and pulled together.	Brows lowered and drawn together and upper eyelids are raised.
AU 1+2+4	AU 1+2+5	AU 1+6	AU 6+7	AU 1+2+5+6+7
				
Brows are pulled together and upward.	Brows and upper eyelids are raised.	Inner portion of brows and cheeks are raised.	Lower eyelids cheeks are raised.	Brows, eyelids, and cheeks are raised.

Figure 2.4: Example of seven upper face action units and how they can be combined to form diverse facial expressions. Image extracted from [106].

The corrugator supercilii and zygomatic are the most explored muscles in the emotion literature [1]. The former is located close to the eyebrow and is related to the eyebrow furrow in negative valence stimuli. While the latter goes from the cheekbone to the corners of the mouth and is correlated with positive valence stimuli [1]. Similarly to the EEG, the EMG has high temporal resolution but low spatial resolution. Therefore, when using the EMG, it can become unclear which muscle activity is being collected since the output can be biased by the activity of neighbour muscles. Likewise, there is a limit to the number of muscles that can be read simultaneously by the EMG, since when several electrodes are placed close together signal interference by cross talk can be observed. Additionally, the EMG requires electrodes to be placed on the face, which is very intrusive and can limit free facial expression [1].

On the whole, the literature has shown that facial expressions (read by the eye or EMG) are associated with primary emotions [84]. However, it's crucial to approach facial expressions with caution. As noted by Mauss and Robinson in [5], facial expressions can be easily feigned and are influenced by a multitude of factors, including gender, culture, expressiveness, and the presence or absence of an audience.

In addition to facial expressions, vocal expressions have also been explored for emotion recognition. For example, higher levels of arousal have been associated with higher fundamental frequencies (e.g. whispering voice is associated with confidentiality, while a strong voice is with anger) [1]. Using basic emotion theory, the literature has identified a set of acoustic characteristics related to a set of basic emotions (Table 2.4 extracted from [72]). For further information, regarding vocal clues for emotion recognition, the reader is referred to [53, 72, 107].

Table 2.4: Voice correlates for basic emotions. A dash indicates that no data is available. Arrows signify changes compared to the baseline, smaller, lower, slower, less, flatter, or narrower ($< \approx$); neutral ($= \approx$); bigger, higher, faster, more, steeper, or broader ($> \approx$); both smaller and equal reported ($< = =$); both larger and equal reported ($> \approx =$); both smaller and bigger have been reported ($< > =$). Table extracted from [72].

Acoustic parameters	Happiness/Elation	Anger/Rage	Sadness	Fear/Panic
<i>Speech rate and fluency</i>				
Number of syllables per second	$> =$	$< >$	$<$	$>$
<i>Voice source—F_0 and prosody</i>				
F_0 mean	$>$	$>$	$<$	$>$
F_0 deviation	$>$	$>$	$<$	$>$
F_0 range	$>$	$>$	$<$	$< >$
F_0 final fall: range and gradient	$>$	$>$	$<$	$< >$
<i>Voice source—vocal effort and type of phonation</i>				
Intensity (dB): mean	$> =$	$>$	$< =$	—
Intensity (dB): deviation	$>$	$>$	$<$	—
Gradient of intensity rising and falling	$> =$	$>$	$<$	—
Relative spectral energy in higher bands	$>$	$>$	$<$	$< >$
Spectral slop	$<$	$<$	$>$	$< >$
Harmonics/noise ratio	$>$	$>$	$<$	$<$
<i>Voice source — glottal waveform</i>				
Excitation strength (EE)	$>$	$> =$	$< =$	$>$
<i>Articulation — speed and precision</i>				
Formants—precision of location	$=$	$>$	$<$	$< =$
Formant bandwidth	—	$<$	$>$	—

One far less explored area to assess emotion is body language. However, studies have shown that emotions such as shame and pride have been associated with contraction and expansion of body postures, respectively [5, 84, 108]. Additional postures such as orientation, degree and direction of trunk lean, head orientation, shoulder orientation, leg orientation, arm openness, and leg openness have been correlated with positive-negative emotions [72].

Generally, body movements are related to basic emotions when they contain a communicative or instrumental function [84]. For example, facial expressions are instrumental in navigating intricate social environments. Similarly, body expressions are associated with emotions like shame, pride, guilt, or embarrassment, which become evident in social contexts, acting as communicative tools to convey feelings to others. On the other hand, basic emotions like anger, fear, disgust, happiness, and sadness — tied closely to individual survival — are mainly represented by distinct facial expressions and generally lack pronounced body expressions [5, 84].

2.2.4 Autonomic Nervous System

The literature on emotion theory (Section 2.1) shows that the ANS is present in most if not all hypotheses. Its level of reported emotion specificity, however, varies.

The ANS takes part in numerous tasks [85]: 1) Regulation – maintaining the body's bodily milieu

by homeostasis; 2) Activation – allocating the body resources in response to stimuli; 3) Coordination – maintaining a continuous bidirectional flow between the body and the brain; and 4) Communication – facilitating non-verbal communication through affective behavioural responses. The ANS multitude of functions and its large number of components add a level of complexity in separating affective-related reactions from other ANS functions.

The ANS is composed of two main components: the Sympathetic Nervous System (SNS), and the Parasympathetic Nervous System (PNS). The SNS is often referred to as the "fight-or-flight" system, while the PNS is denoted as the "rest-and-digest" system. Although for the majority of its functions, the SNS is indeed an activator and the PNS a deactivator, that is not observed across all of their functions. The SNS is responsible for the increase in HR through vasoconstriction, bronchial tube dilation, muscle contraction, pupil dilation, decrease in stomach movement and secretions, decrease in saliva production, and, lastly, the release of adrenaline. On the other hand, the PNS is responsible for slowing down the HR, vasodilation, increase in the digestive system activity, bronchial tube constriction, muscle relaxation, pupils constriction, increase in stomach movement and secretions, increase in saliva production, and lastly, the increase in urinary output [9, 85]. These functions illustrate the "fine-tuned" work by the ANS in its multitude of functions.

As the Section 2.1 on emotion theory showed, there is a heavy debate in the field about whether emotion and its components have differentiated patterns. This discussion is explored through the ANS measures by analysing ANS coherence and specificity. Coherence refers to both within the ANS (i.e. cardiac, vascular, electrodermal response) and with the other components of emotional response (e.g. facial expressions, or CNS). Specificity refers to the Essentialism view, i.e. whether a pattern is unique to each emotion category.

Regarding coherence, Mauss *et al.* [109] analyzed the within-ANS response, by measuring subjective experience (through continuous self-reports), facial behaviour and peripheral response (through EDA, cardiovascular activation and somatic activity), while the subjects watched a 5-minute film. The experimental results showed that the within-ANS was high for amusement, with elevated coherence for self-reports and facial expressions, increasing with the amusement experience intensity. However the same was not observed for sadness intensity.

For specificity, Levenson [85] delineated structural and functional features of the ANS that could account for a patterned activation, namely: 1) The existence of different types of receptors in the SNS, each with different sensitivities and body distributions; and 2) The existence of two branches arousal-related – with separated structure, regions of activation and receptors, and acts in combination on common areas (towards a more complex ANS regulation).

Moreover, Levenson [85] defended a few aspects that must be met to correctly evaluate the ANS specificity, among these are: 1) Verifying that the emotion was elicited (self-report); 2) Attaining a high-

intensity emotional state (naturalistic scenario); 3) Performing a broad assessment of the ANS (rely on more than one system); 4) Extracting comprehensive features that characterize the ANS; and 5) Relying on a correct timing that accounts for the ANS systems relation to emotion. The lack of compliance with these criteria has led to emotion specificity being a heavily debated topic in the literature, with several works in support [5, 85, 110, 111, 112] and opposition [113, 114, 115, 116].

The primary systems within the ANS associated in the literature with emotion are the cardiovascular and integumentary. The signals related to such systems offer the benefit of being measured from discreet body locations, allowing for data collection in naturalistic, out-of-lab settings. Thus, more likely to capture high-intensity emotional responses. Within these systems prevalent sensors are the EDA, governed by the eccrine sweat glands, and the PPG, a non-invasive approach to monitor blood volume changes and HR. The two modalities combined offer a comprehensive insight into the SNS and PNS dynamics. A more detailed description of these systems is presented next.

Electrodermal Activity

The EDA (often also referred to as Galvanic Skin Response (GSR)) denotes the "*changes in the electrical activity of palmar and plantar skin*" [117, 118]. Changes in the electrical characteristics of the skin are read by changes in the skin conductance or its inverse, the resistance. When the sweat glands are activated by the SNS, there is an increment of sweat, which is mostly water with salts. Since sweat contains a high quantity of electrolytes, it will increase the conductance of the skin [118]. Although we often sweat due to intense physical activity, or due to the outside temperature through body-regulation mechanisms, on the palmar (hand palms) and plantar (foot soles), we observe emotional sweating: sweat activated by psychological and emotion-related events [119]. This is due to the hands and palms being body areas with a high concentration of eccrine glands (sweat glands related to thermoregulation and ANS changes), leading to pronounced changes in skin resistance which can be used to detect emotion-related body changes.

The EDA can be decomposed in the Electrodermal Level (EDL) and Electrodermal Response (EDR) components (Figure 2.5). The EDL is referred to as the tonic component, while the EDR is referred to as the phasic component that emerges from the EDL during an emotional response. When the change in the EDA amplitude is due to a known stimulus, the EDA response is denoted as Event-related EDR (ER.EDR). While, when the change in the EDR can not be related to any emotional stimulus or artefacts (e.g. movement or change in breathing), it is denoted as non-specific EDR (NS.EDR) [118].

As depicted in Figure 2.6 a typical EDA signal is characterized by an abrupt rise from the baseline level (upon emotional stimuli), a peak (when the sweat glands are emptied), and a slow decline to the baseline level. Typical features extracted from the EDA signal are shown in Table 2.5.

Eccrine sweat glands are enervated by the SNS but not by the PNS. As such, the analysis of the

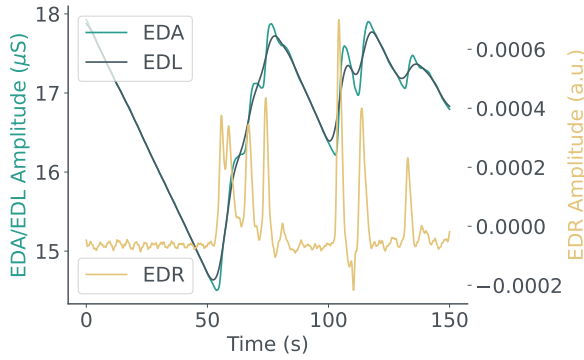


Figure 2.5: EDA signal decomposed in its two components: EDL and EDR.

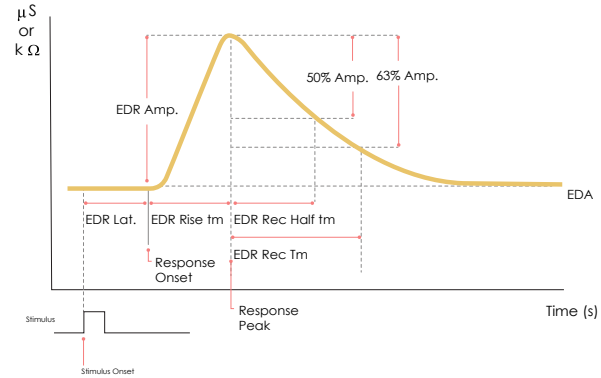


Figure 2.6: Prototypical EDA waveform and extracted features.

Table 2.5: Typical features extracted from the EDA signal and its expected times/amplitudes [117, 119].

Metric	Time (s)	Amplitude (μS)
Latency Time	1 - 5	
Rise Time	0.5 - 5	
Recover Time (ER.EDR)	4 - 6	
Recover Time (NS.EDR)	10 - 30	
ER.EDR minimum amplitude		0.01 - 0.05

skin conductance decomposes the ANS complexity by offering an unbiased view of the SNS (either emotional or physiological functions) [120].

Moreover, the EDA is the most common measure for the arousal dimension, with EDL and NS.EDR frequency being predominant extracted tonic metrics [118]. Lang *et al.* [81] observed the link between the ANS and dimensional/core theories of emotion experience by obtaining a correlation of 0.81 with EDR log magnitude. In the context of discrete-emotion representation, the meta-analysis review by Kreibig [110], surveyed 134 studies to identify autonomic signatures. The results showed that EDL is the variable most often reported, followed by EDR rate and EDR amplitude. An alternative review by Mauss and Robinson [5] reported that EDL or short-duration EDA responses are associated with SNS activity. Table 2.6 shows the modal EDA direction changes reported in the Kreibig comprehensive meta-analysis review [110]. Overall, the literature shows that an increase in the EDA was identified for most of the discrete emotions, related to cognitive or affective-induced motor preparation. On the other hand, a decrease in EDA was observed for non-crying sadness, acute sadness, contentment and relief, related to passivity and a decrease in motor preparation [110].

Cardiovascular Activity

The heart is influenced by both the SNS and PNS, which can work together or individually to increase or slow down its activity. Common cardiovascular measurement methods include the PPG and Electrocardiography (ECG) sensors. From these sensors, features relevant to emotion recognition can

Table 2.6: Reported EDA modal responses across discrete emotions. The response direction identified at least in three studies is reported, except for the arrows in parentheses, reported in fewer than three studies. Arrows denote changes in activation levels from the baseline: increase (↑); decrease (↓); no change (-); or diverse responses across studies (↑↓). Table adapted from the work [110].

	Anger	Anxiety	Disgust Contamination	Disgust Mutilation	Embarrassment	Fear	Fear imminent threat	Sadness Crying	Sadness Noncrying	Sadness Anticipatory	Sadness Acute	Affection	Amusement	Contentment	Happiness	Joy	Affect Pleasure Visual	Affect Pleasure Imagery	Pride	Relief	Surprise	Suspense
EDR	↑	↑	↑	↑		↑	↑				↓		↑	(...)	↑		↑			↓		
NS.EDR	↑	↑	↑	↑		↑	↑	↑	↓	↑	↓	(↑)	↑		↑		↑			(↓)		(↑)
EDL	↑	↑	↑	↑	(↑)	↑	↓	↑	↓	↑	↓		↑	↓	↑...	-	↑		↑	↓	(↑)	(↑)

be extracted, namely HR, blood pressure, total peripheral resistance, cardiac output, pre-ejection period, or Heart Rate Variability (HRV) [5]. The pre-ejection period has shown to be associated with the SNS, HRV with the PNS, while HR and blood pressure, with a combination of both [5].

The ECG measures the heart’s electrical activity by detecting potential differences propagated to the skin’s surface, which result from the contraction and relaxation of the cardiac muscle due to electrical stimulation. The heart’s contraction and relaxation are the result of three main components: (a) The action of the sinoatrial node, localised in the right atrium at the superior vena cava, which receives inputs from the branches of the ANS to beat at a resting rate frequency (100 to 120 beats per minute) [85]; (b) The action of PNS fibres modulated by the vagal nerve, to slow down the HR to approximately 70 bpm; (c) The SNS fibres modulated by the post-ganglionic fibre to increase the HR during an emotional episode or non-emotional ANS modulation [85]. The complementary action of the two branches of the ANS system in the HR is known as sympathovagal balance.

In this thesis, given the focus on naturalistic data collection, the PPG sensor by facilitating data collection in unobtrusive locations, and the metrics that can be extracted from both ECG and PPG signals (e.g. HR) will be the focus.

The PPG measures peripheral tissue’s blood volume through the absorbance of the light. The PPG sensor consists of a light source (light-emitting diode (LED)) and a light detector (photodiode). The detector can be placed next to the light source to read the back-scattered light (Reflective PPG in Figure 2.7) or across the finger (Transmissive PPG). The most usual PPG setup in wearable systems is the reflective mode, which will be the one used in this work, using a green light. Greenlight is preferred over infrared and red wavelengths due to higher skin penetration and Hb absorptivity [121, 122]. Figure 2.7 displays a prototypical PPG waveform.

When light is directed into the skin, it is absorbed by blood, tissue and others. The more light is absorbed, the less it reaches the sensor detector through light scattering or reflection. Moreover, the

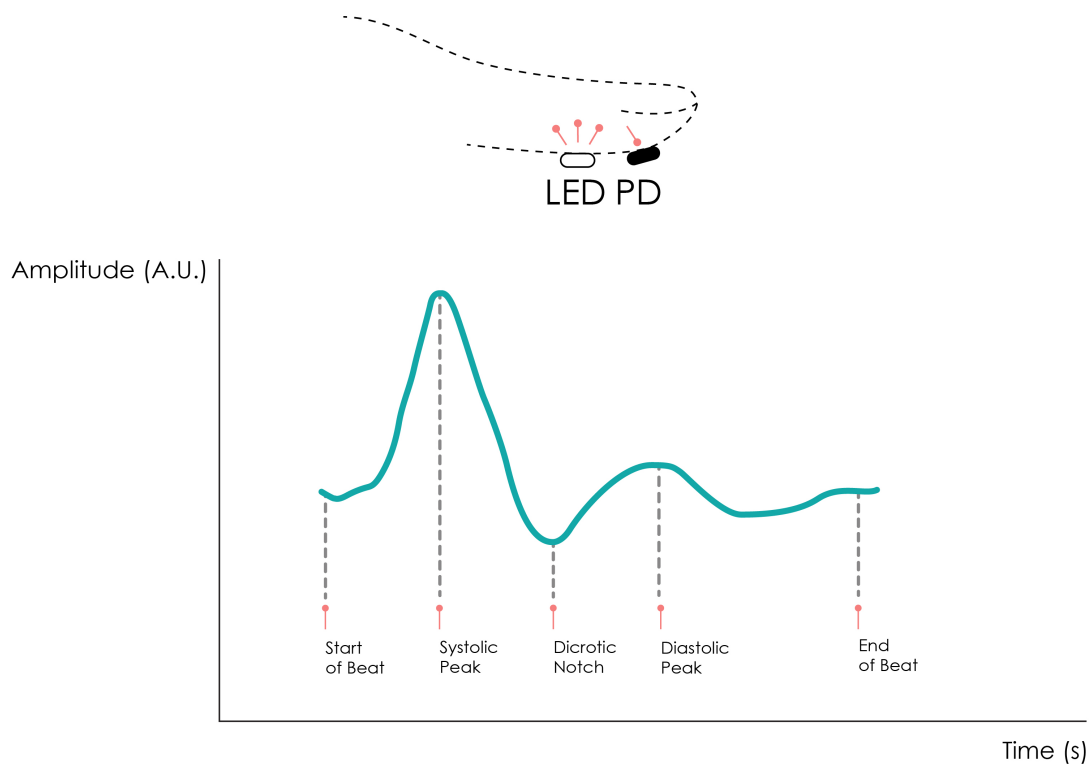


Figure 2.7: Prototypical PPG waveform. Image extracted from [13].

tissue absorption changes according to the Hb configuration (oxygenated versus deoxygenated) and the arterial blood volume which increases during systole and decreases during diastole. During systole, the arterial blood volume is maximum and the PPG signal reaches its peak (Systolic peak in Figure 2.7). The opposite is observed during the diastole where the PPG reaches its minimum value (Diastolic Notch in Figure 2.7). During the diastole, the PPG will reach its baseline component, reading the absorption from the blood volume in the veins, capillaries, volumes of skin, fat, and bone, among others. Thus, the PPG is often decomposed into two components: a pulsatile waveform from arterial blood and an offset component from respiration, venous, capillary blood, and stationary tissues [123].

Table 2.7 shows the modal cardiovascular changes reported by Kreibig [110]. Overall, the HR was identified to increase in negative (anger, anxiety, contamination-related disgust, embarrassment, fear, crying-sadness), positive (imagined anticipatory pleasure, happiness, joy) emotions and surprise. On the other hand, HR decreases for emotions involving passivity (mutilation-related disgust, imminent-thread fear, non-crying and acute sadness, affection, contentment, visual anticipatory pleasure and suspense).

Regarding positive emotions, HRV was reported to differentiate positive emotions, increasing for amusement and joy, while it decreases for happiness. An increase in HRV was observed in contamination-related disgust, and the positive emotions: amusement and joy. The results from cardiovascular activity support the hypothesis of PNS enabling both positive and negative-related emotions [110].

Table 2.7: Cardiovascular modal responses across discrete emotions. Arrows denote changes from the baseline: increased (↑); decreased (↓); no change (-); or both increases and decreases across different studies (↑↓). The response direction identified at least in three studies is reported, except for the arrows in parentheses, reported in fewer than three studies. Figure extracted from [110].

	Anger	Anxiety	Disgust Contamination	Disgust Mutilation	Embarrassment	Fear	Fear imminent threat	Sadness Crying	Sadness Noncrying	Sadness Anticipatory	Sadness Acute	Affection	Amusement	Contentment	Happiness	Joy	Affect Pleasure Visual	Affect Pleasure Imagery	Pride	Relief	Surprise	Suspense
HR	↑	↑	↑-	↓	↑	↑	↓	↑	↓	↑	↓	↓	↑↓	↓	↑	↑	↓	↑	↑↓	↑-	↑	(↓)
HRV	↓	↓	↑	-	(↓)	↓	(↓)	-	↓	(↓)	↑	↑	↑	↑↓	↓	(↑)	(↑)	↑	(↓)	↑-	↑	(↓)
LF		(↑)				(↓)									-							
LF/HF	(↑)	(↑)		(↓)									(-)									

Other Measures

In addition to the cardiovascular and electrodermal activity, additional data with information for emotion specificity has been identified in the respiratory system. The respiratory system has a predominant role in preparing the organism for contexts that require motor changes (e.g. fight-or-flight) [85]. Among respiration measures, the respiration rate is the most often extracted characteristic, followed by the respiratory period, respiratory depth, tidal volume, duty cycle, and respiratory variability [110]. Fast deep breathing has been found to be correlated to non-crying sadness (to suppress active crying), slow deep breathing to relief, and shallow breathing to anxiety, disgust, certain types of sadness, and anticipatory pleasure [110]. For a deeper analysis on ANS-related respiration changes in discrete emotions, the authors refer the reader to the work [110]. A downside to the respiration sensor is that it requires the positioning of a respiration belt on the chest area, which can be uncomfortable.

Lastly, in our daily lives, we often refer to expressions such as "gut feeling" or "feeling sick to my stomach" to denote negative feelings. The stomach and intestinal activity are activated by the PNS, thus could contain meaningful information for emotion ANS specificity. However, the gastrointestinal system has received limited attention in the literature and will not be the focus of this thesis.

2.2.5 Discussion

In this section, the multi-dimensional components of emotions were analysed. For the scope of this thesis, the goal of which is to bring emotion recognition to the real world, imaging and physiological measures techniques related to the CNS have severe limitations. The EEG requires the placement of a high number of electrodes and conductive gel on the hair. Similarly, imaging techniques can only be performed in specialized facilities and by technical individuals. Moreover, the data collection machines present numerous disadvantages as they are large in volume, have an elevated economic cost, require

the presence of a trained technician, and require the individual to remain motionless, among others.

An alternative method is to analyze bodily expressions. However, studies have found that smiling when people are happy tends to happen more frequently in the presence of an audience [5, 124]. There is an open debate on whether facial and body expressions reflect internal states or are rather deliberate communications [125].

An additional component of emotion is the ANS assessed through physiological measures. Thanks to the advance of electronics, physiological signals can be measured through miniaturized devices in unobtrusive locations such as hands or wrists, and are already integrated into numerous smartwatches or fitness bands. This makes physiological signals like EDA and PPG ideal for collecting large amounts of data in the real world effortlessly and with minimal disturbance to the user.

On the whole, contradictory conclusions are reported, with diverse autonomic responses even for instances of the same emotion category. However, it is unclear if that is a product by design, i.e. emotions simply do not follow a discrete pattern, or rather, a result of poor experimental methods, weak emotion elicitation, or insufficient statistic tools to analyze the complexity of the signals [74].

The literature [115, 6] reports that in-lab responses can result in stereotype emotions, differing from real-world responses, or be mild in intensity, the reported responses from the state of the art, being based on in-lab experiments can differ when explored in a naturalistic scenario. As Quigley and Barret [116] denote, in-lab experiments are likely unrepresentative of the broad scope of emotional states experienced in the real world, and are most likely of small to moderate intensity, becoming easily masked by other ANS-mediated bodily changes.

In the future chapters of this work, namely Chapters 3 to 5, building upon these conclusions, physiological data for emotion recognition is explored, namely through EDA and PPG data. By allowing the unobtrusive data collection, these pave the way for a more naturalistic data collection, that better approximates the emotion elicitation and the obtained data to the real world as will be explored in Chapters 6 and 7.

2.3 Datasets

There are multiple publicly available datasets for emotion recognition based on physiological data. They allow researchers to develop and compare different algorithms on the same data, without the experimental effort of collecting it. Table 2.8 displays prevalent datasets in the emotion recognition state of the art.

- **Setting:** The overview of the literature shows that most datasets for emotion recognition acquire data in controlled, in-lab scenarios. This setup results in the use of small video clips for emotion elicitation, except for a few datasets that although contain data collected in laboratory settings, are

Table 2.8: Literature datasets for emotion recognition using physiological data. Nomenclature: #S (Number of Subjects); Set. (Setting); L (lab), W (in-the-wild); Application: ER (Emotion Recognition); SR (Stress Recognition); Annotation space: A (Arousal); V (Valence); D (Dominance); BET (Basic Emotions); RESP (Respiration); SKT (Skin Temperature). Table adapted from [10].

Name	#S	Stimuli	Set.	Area	Annotation	Data
DEAP [89]	32	40 music videos (1min)	L	ER	A, V, D, liking, familiarity	ECG, EDA , EEG, EMG, EOG, RESP, SKT, face video
MAHNOB [126]	27	20 film clips (35-117s)	L	ER, implicit tagging	A, V, D, predictability, BET	ECG, EDA , EEG, RESP, SKT, face and body video, eye gaze tracker, audio
ASCERTAIN [127]	58	36 movie clips (51-128s)	L	ER, Implicit personality	A, V, liking, engagement, familiarity, Big Five	ECG, EDA , EEG, facial video
Eight-Emotion [128]	1	8 posed emotions	L	ER	Neutral, anger, hate, grief, joy, platonic love, romantic love, reverence	ECG, EDA , EMG, RESP
EMDB [129]	32	52 non-auditory film clips (40s)	L	ER w/out auditory content	A, V, D	EDA , HR
AMIGOS [87]	40	16 250s videos; 4 videos (14min) alone and in group	L	ER, personality traits and mood on individuals and groups	Big-Five personality traits, PANAS, A, V, D, liking, familiarity and BET	Audio, visual, depth, EEG, GSR and ECG
DECAF [130]	30	40 music video (1min DEAP), 36 movie clips	L	ER	A, V, D	ECG, EMG, EOG, MEG, near-infrared face video
WESAD [88]	15	baseline (20mins), video clips (392s), TSST (5min), meditation (7mins)	L	ER, SR	SAM, PANAS, SSSQ and STAI	ACC, ECG, EDA , EMG, RESP, SKT; ACC, EDA , PPG , SKT
CASE [131]	30	Video clips (101-197s)	L	ER	A, V	ECG, PPG , EMG (3x), EDA , RESP, SKT
CLAS [132]	62	Math and logic problems, Stroop test, IAPS, multimedia clips (30min)	L	ER, SR	A, V	ECG, PPG , EDA , ACC
PAFEW [133]	24	480-6040ms short videos	L	ER	7 BET	EDA , PPG , SKT, pupil
CEAP [134]	32	360VR videos (1min)	L	ER	A, V	Acc, EDA , SKT, Blood Volume Pulse (BVP) , HR
ITMDER [135]	18	7 VR videos (43 to 210 s)	L	ER	A, V	ECG, EDA , RESP, PPG
BIRAFFE2 [136]	102	Video games (5 min each) + 6s IAPS, IADS	WL	ER	A, V	ECG, EDA , Acc, Gyr
K-EmoCon [137]	32	Debate (10min)	WL	ER	Retrospective A, V, 18 BET	EEG, ECG, PPG , EDA , SKT
COGNIMUSE [138]	7	Movies	WL	ER	Feeltrace A, V	None
LIRIS-ACCEDE [139]	10	30 movies	WL	ER	GTrace A, V	EDA , motion, SKT
DAPPER [140]	88	5 days of daily living	W	ER, SR	ESM, DRM	HR , EDA , ACC

closer to naturalist scenarios, i.e. by using long-duration content (>10 minutes) as the elicitation method, namely AMIGOS [87], COGNIMUSE [138], and Continuous LIRIS-ACCEDE [139]. Alternatively, the DAPPER dataset [140] and K-EmoCon [137] collect data in naturalistic scenarios. The former follows a lifelog paradigm, with physiological data (HR, EDA, and accelerometer) collected during the volunteers daily living for five days. The data was annotated using the experiment sampling technique and day reconstruction method. The experiment sampling was performed 6 times a day, asking for the momentary emotional annotation on the volunteers' smartphone; and the day reconstruction questionnaire was performed at the end of the day asking the volunteers to recall and annotate the major emotional and behavioural episodes throughout the day. The K-EmoCon dataset [137] contains physiological data (PPG, EDA, HR, EEG) collected across naturalistic conversations, namely paired debates on social issues. The data was annotated retrospectively by the participants, their debate partners and themselves by watching their recorded facial expressions and upper body data.

- **Stimuli:** As previously noted, most of the in-lab datasets rely on small video clips to elicit emotions (< 5 minutes), with a few exceptions. The AMIGOS [87] dataset contains data from 16 short emotional videos, and 4 long videos (alone and in group settings) with \approx 14 to 24 minutes, including EDA, ECG, and EEG data. The long video data was segmented in 20-second windows and annotated using the individuals' facial expressions by three external annotators in the continuous

arousal and valence dimensions using SAM. The COGNIMUSE [138] dataset contains data from 7 Hollywood movies annotated in arousal and valence. Two types of annotations are provided: experienced self-reports by the volunteers; and, intended emotions annotated by experts. Both types of annotation are performed using the FEELTRACE platform [141], and no physiological data is acquired. Continuous LIRIS-ACCEDE [139] contains data from 30 movies (M=884s, STD=766s) annotated on valence or arousal axes by a modified GTrace platform [142] incorporating SAM and a joystick. During video visualisation, EDA, finger motion, and skin temperature were recorded. The K-EmoCon dataset [137] as previously mentioned, contains data on a debate on social issues, and DAPPER [140] contains data from the daily living of the volunteers.

- **Annotation Dimensions:** On the whole, the valence and arousal dimensions are the preferred choices for the subjective reporting of emotional experience. The use of discrete emotions is also observed [126, 127, 133]
- **Data:** Across the datasets, information regarding cardiovascular data (either through ECG, PPG or HR) and EDA is predominant.

The review of the state of the art shows that, overall, the available datasets for emotion recognition based on physiological data use small video clips (excerpts extracted from longer videos) as the emotion elicitation method and are acquired in the lab. The use of a few seconds/minute clips detaches the elicitation from a naturalistic elicitation, where longer-duration content such as the entire movie/TV show is watched. Media content is usually devised taking into account an emotional timeline that builds to the elicitation of a certain emotion, which the use of small video clips limits.

Lastly, the literature has focused on the analysis of individual emotional responses. Ignoring group-level factors that are often present in social contexts in the real world. An exception is the AMIGOS [87] and K-EmoCon [137] datasets, which contain data from four subjects watching long videos and from dyad conversations, respectively. Moreover, the authors of AMIGOS [87] analysed the impact of the group context and identified that social context affects the valence and arousal self-reports of the participants, attaining for the long-videos overall higher levels of emotion intensity. Leading to the conclusion that a further line of research should collect data in longer-term content and explore group data for a more holistic overview of individual and group emotions.

2.4 Group Emotions

This thesis started on the premise that emotions are a powerful phenomenon capable of influencing our thoughts and behaviours. However, emotions influence not just ourselves but the emotions and consequent behaviours of others, just as we are influenced by the emotions and behaviours of those

around us [143]. This interconnectedness underscores the significance of emotions within the broader constructs of groups and societies. To illustrate this concept:

"Imagine attending a concert for a world-famous band you adore. As the lights dim and the opening notes play, you can feel the excitement radiating from thousands of fans. When the band plays an iconic song or a ballad, the whole stadium might sing along, the collective voices creating a powerful and emotional resonance. When the audience waves their hands or lighters, or when everyone jumps in unison, individual attendees often feel an overwhelming sense of unity and shared emotion" – Example of group emotions by Chat-GPT.¹

In moments like these, just as it happens with religious ceremonies, sports gatherings or crowd manifestations, the group emotions can be palpable. If the majority of the crowd feels a particular emotion (e.g., euphoria, nostalgia, excitement), it's easy for individuals to get swept up in the collective atmosphere.

2.4.1 Groups

The sense of shared emotion serves as a foundation for group cohesion and identity. A group emerges when individuals identify themselves as a group member and the group becomes a meaningful aspect of its identification. Menges and Kilduff [144] define a group as *"a number of people who are connected by some shared activity, interest, or quality"*. Under events that are relevant to the group identity, people tend to follow the group beliefs, attitudes and even behaviours. This can arise from small face-to-face meetings (dyads) to large collectives (without face-to-face interaction such as demographic identities) (Table 2.9).

Table 2.9: Groups compositions across diverse sizes [144].

Groups	Examples
Dyads	Peer Relationships, Supervisor–sub-Ordinate Pairings
Small Groups	Work Groups, Teams
Mid-size Groups	Branches, Departments, Organizations
Large Groups	Industries, Demographic Groups

2.4.2 Collective Emotions

Early research on group emotions tended to adopt a somewhat pessimistic perspective. McDougall [145] described group membership as akin to being *"carried away by forces"*, over which one felt *"powerless to control"* [125]. Contrasting with this view, Durkheim [146] perceived group emotions in a more

positive light, suggesting that in large gatherings where participants share a common purpose or emotion, the sense of belonging and connection to the group could foster a "collective effervescence" – an empowering phenomenon that can reinforce shared ideas, values, and actions [147].

As the understanding of the field evolved, such that emotions were not solely experienced at the individual level but also manifested as a higher-level process (the group emotions level), the field of collective emotions recognized its intricate and empowering nature, leading to a surge in interest.

Kelly and Barsade [148] reported that at the individual level, individuals are composed of affective-related characteristics, namely: moods, emotions, affective dispositions, emotional intelligence and sentiments. When an individual joins a group, they bring their unique emotional characteristics with them. These characteristics interact with the other group members' emotional characteristics and are influenced by contextual factors (e.g. cultural norms, and history of the group). These factors dynamically shape what becomes the group's emotional atmosphere. Menges and Kilduff [144] further articulate this construct, defining group emotions as "feelings that emerge from, or in, groups". Moreover, the authors divide group emotions into group-shared emotions and group-based emotions. Group-shared are synchronous, interactive emotions that individuals collectively experience in groups, while group-based emotions are asynchronous, non-interactive emotions elicited towards the group identity [144].

In a similar view, Goldenberg *et al.* [149] propose a broader delineation of emotions. They differentiate emotions into individual, group-based, and collective emotions (Figure 2.8), where individual emotions are composed of individual-level emotions and group-based emotions. Individual-level responses are the multi-component synchronized responses upon relevant stimuli detailed in Section 2.1. Group-based emotions are individual emotions (occurring at the individual level) where the stimulus is directed not towards the individual, but to the group identity (e.g. group-based joy when our country excels in the Olympics). Lastly, collective emotions are the macro-phenomena of many individuals experiencing emotions together (e.g. joy from being in a crowded stadium supporting your team after a win). Individual emotions turn into collective emotions by interaction-led emotion dynamics (emotion transfers) which often result in a shared identity or a common goal.

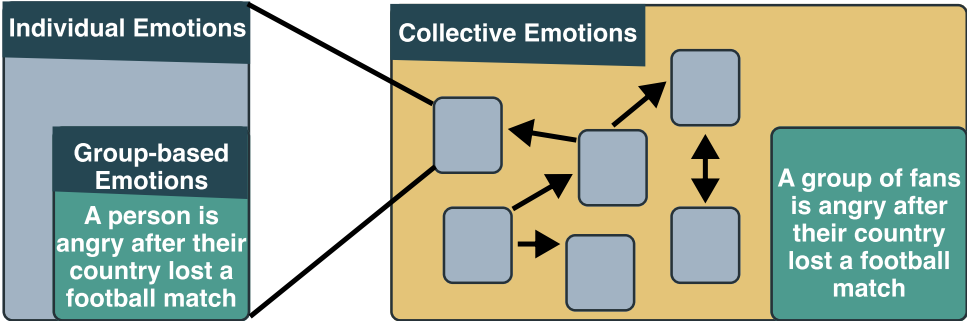


Figure 2.8: Illustration on how individual emotions, at both personal and group levels, interact and transfer between individuals to culminate in collective emotional experiences. Image adapted from [149].

The distinctions between Menges and Kilduff's concept of shared emotions [144] and Goldenberg's notion of collective emotions [149] are ambiguous. Menges and Kilduff emphasize interactivity, with emotional transfers from direct interactions among small to medium group members, potentially leading to emotional convergence or divergence. In contrast, Goldenberg focuses on the shared emotional state and the macro-level implications within larger groups [149].

2.4.3 Emotion Transmission in Groups

Expanding on the concept of group emotions and how they manifest, Barsade and Gibson [150] characterize the processes that lead to group emotions in two components: a Bottom-Up component (affective compositional effects), where the group affect emerges from the combined individual-level affective factors of its group members [151]; and a Top-Down component (affective context), where the group environment influences the emotions of the individuals within. Kelly and Barsade [148] describe the two components as combined forces that act towards the creation of the group emotion.

In the Bottom-Up process, emotions at the individual level can propagate among individuals through two mechanisms:

Explicit (Conscious) Processes

In social contexts, individuals consciously adjust their emotions in response to environmental cues. This deliberate affective regulation is denoted as intentional affective induction, where individuals either attempt to manipulate or become influenced by the emotions of their peers (e.g. a project team, members might display exaggerated enthusiasm for an idea, aiming to gain consensus and approval) [148]. The affective induction can result in both concordant and discordant emotional outcomes, with the latter occurring less frequently. A parallel phenomenon is affective impression management, wherein individuals feign emotions to align more closely with group norms.

Implicit (Sub-conscious) Processes

In the domain of innate mechanisms, emotional exchanges occur automatically among individuals, encompassing phenomena such as [148]:

- **Emotion Contagion:** Emotions from one individual are spread to those nearby. The literature reports an innate and unconscious inclination among individuals to mimic and synchronize with others' facial expressions, vocalizations, postures, and movements, leading to emotional convergence [148, 152]. Such shifts in bodily behaviour can evoke corresponding emotional states, (e.g. mimicking a smile can induce feelings of joy [148]). Several factors can impact emotion contagion,

such as the strength of the relationship, the body's expressiveness, or the subject overall mood [148].

- **Vicarious Affect:** Emotion induction by observing or understanding another person's emotions, e.g. empathy, where we feel "*another's feelings by placing oneself psychologically in that person's circumstances*" [153].
- **Behaviour Entrainment (Mimicry):** Emotion induction through synchrony, i.e. the subject adjusts their behaviour to mimic and synchronize with another. Mimicry can lead to emotional convergence, as detailed by Hatfield *et al.* in [154] "*the tendency to automatically mimic and synchronize facial expressions, vocalizations, postures, and movements with those of another person and, consequently, to converge emotionally*". Behavioural synchrony is likely to lead to positive affect, denoting satisfaction with the interaction or the liking of the group.

As a result of the aforementioned processes, the group emotion will be created. When the group has a homogeneous group affect it is said to have a Group Affective Tone [155].

Menges and Kilduff [144] further explain the constructs that lead to the development of Bottom-Up mechanisms in collective emotions:

- **Inclination:** Derives from the process wherein individuals with similar affective dispositions naturally tend to similar emotions, achieving a homogeneous emotional state.
- **Interaction:** Derives from social interactions as the basis for emotion contagion (previously detailed) and sensemaking, related to experiencing the same stimulus and appraising it in an equal way (parallel emotion elicitation), leading to a similar emotion.
- **Institutionalization:** Derives from the need of the individual to blend with the institutionalized norms, rituals and routines, with expected emotional scripts which can lead to emotional convergence (e.g. collective effervescence [146]).
- **Identification:** Derives from the subject seeing themselves as part of the group and feeling emotions towards the group identity (e.g. group-based emotions [149]).

In addition to Bottom-Up processes, the individual's affective state, and by extension the emotional make-up of the group, can be influenced by Top-Down processes, namely the group and local norms, as well as the group history, to enhance or suppress an individual's emotions.

On the other hand, in addition to emotion-related processes, non-affective processes [148] such as the inter-group context (relation with other groups), the comfort of the physical context (e.g. the ambient temperature), or technological interface (communication in remote which could impact non-verbal clues) can impact emotional dynamics.

As a result of the Bottom-Up transfers of emotions and the Top-Down affective context, the individual-level emotions will be manipulated into a dynamic group emotion. Goldenberg *et al.* [149] attempt to characterize the dynamic emotional transference in groups using three dimensions: quality, magnitude and time course.

When people interact in groups, the group individual's emotion variance can decrease and converge – Emotion Contagion. Or, in opposition, emotional dynamics can lead to the polarisation of group members towards each other or to the group, and lead to an increase in the group-level variance, with a change in the type of emotion and an increase in intensity. Similarly, upon changes in the emotion type, the emotional state of one person can serve as elicitation for the emotional state of another. Secondly, there can be changes in the magnitude of the emotional responses. In the presence of other peers, emotions tend to show higher emotion intensity, through processes such as emotion contagion/collective effervescence. Lastly, collective emotions can provoke changes in the timeline of individual-level affective experiences: At the individual level, emotions are intense but short-lasting events. However, in groups, people tend to activate each other in a phenomenon denoted as emotional cascades, maintaining the intensity of emotional responses at a high level by emotion contagion for a longer period [149].

2.4.4 Discussion

The exploration of group emotions showed that in groups, individuals will undertake emotional transfers from Bottom-Up transfers and the impact of a Top-Down affective context, towards the macro-level phenomenon of Collective Emotions. Moreover, predominant Bottom-Up processes act as channels for the dissemination and alignment of emotions across group members. These processes, namely emotional contagion have been reported across various groups and activities from employees and customers [156], leader-followers [157], friend-to-friend in both online and in person [158], or in smaller groups [144, 159].

Such dynamics suggest that discerning the emotional state of one member can offer invaluable insights into the collective, and in the context of emotional convergence, potentially predict the emotional states of other participants. Taking the componential view of emotion, these homogeneous emotional states can be analyzed through physiological synchrony, one of the components of emotional responses. This group approach contrasts with traditional algorithms for emotion recognition based on physiological signals and data collection tools which have largely concentrated on singular analyses of individual physiological responses, thus paving the way for augmenting and refining emotion recognition technologies.

This thesis builds upon group emotions, first to explore physiological synchrony as a tool for group emotion recognition in publicly available datasets in Chapter 4, secondly to develop the required infrastructure to capture group physiological data in diverse group settings in Chapter 5, annotate it Chapter 6, and lastly to create a dataset containing physiological data collected in groups in Chapter 7.

Chapter 3

Benchmarking Emotion Recognition

In the last chapter, an analysis was conducted on the nature of emotions and the methods by which emotion can be measured, concluding that physiological signals are the most beneficial for real-life data collection.

This chapter delves into emotion recognition through physiological signals through *Obj 1* on the *Development of Affective Computing Algorithms* and explores RQ 1.1 to 1.3 by examining the current state of the field on the algorithms employed, extracted features, validation techniques, among others. This is followed by the experimental validation of these methods and the establishment of a baseline performance across public datasets.

The contents within this chapter were adapted, with permission, from:

- P. Bota, C. Wang, A. L. N. Fred, and H. Plácido da Silva, “A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals,” *IEEE Access*, vol. 7, no. 1, pp. 140990–141020, 2019
- P. Bota, C. Wang, A. Fred, and H. Silva, “Emotion assessment using feature fusion and decision fusion classification based on physiological data: Are we there yet?,” *Sensors*, vol. 20, no. 17, p. 4723, 2020

Contents

3.1	Introduction	42
3.2	Background	43
3.2.1	State of The Art	43
3.2.2	Machine Learning	54
3.3	Methods	55
3.3.1	Data	55
3.3.2	Signal Pre-Processing	56

3.3.3	Feature Extraction and Selection	56
3.3.4	Classification	57
3.3.5	Evaluation	57
3.4	Results	58
3.4.1	Single Modality Models	58
3.4.2	Multi-Modality Models	59
3.5	Discussion	63
3.6	Conclusion	63

3.1 Introduction

In 1995, Rosalind Picard [2] created the field denoted as affective computing, dedicated to the study of affective phenomena by computational means. The field has since then grown motivated by its potential applications in healthcare, education, entertainment, and marketing, among others. For example in healthcare, wellness monitoring holds the promise of identifying the causes of stress, anxiety, depression or chronic diseases; In education, adaptive learning could be used to adjust the content delivery rate and several iterations according to the user enthusiasm and frustration level; Or in entertainment, emotion recognition could be used to adapt narratives to emotional responses, and enhance live performances through feedback loops between actors, audience, and technical elements. The potential applications are vast, however, the majority of these applications are still in their infancy as the predominant focus in the field is on the development of emotion recognition algorithms and their underlying structure (i.e. creating reliable data collection devices, annotation tools and datasets). To develop these systems, several critical questions remain, including the choice of classifiers, sensor modalities, feature selection, window sample sizes, validation metrics, expected accuracy ranges, and others. To address these issues, an exploratory evaluation was conducted, focusing on:

RQ 1.1: *What feature set/machine learning algorithm should be applied for emotion recognition?*

RQ 1.2: *What performance can be achieved by predominant datasets in the literature?*

RQ 1.3: *What is the best method to deal with multi-modal data for emotion classification?*

This chapter addresses these pivotal questions through a comprehensive literature review on emotion recognition, specifically using physiological data. The review is then complemented by a benchmarking analysis of the performance of the main emotion recognition classifiers identified in the literature on ECG, EDA, respiration, and PPG data. The analysis relies on five publicly available datasets: IT Multimodal Dataset for Emotion Recognition (ITMDER) [135]; Multimodal Dataset for Wearable Stress

and Affect Detection (WESAD) [88]; Dataset for Emotion Analysis using Physiological Signals (DEAP) [89]; Multimodal dataset for Affect Recognition and Implicit Tagging (MAHNOB) [126]; and Eight-Emotion Sentics Data (EESD) [128], offering insights into the current capabilities and limitations of emotion recognition classification.

3.2 Background

This section surveyed over 50 papers on the field (see Figure 3.1) to provide a comprehensive overview of the state-of-the-art in emotion recognition. The papers were selected based on their relevance to the field, the quality of the research, and the availability of the full text.

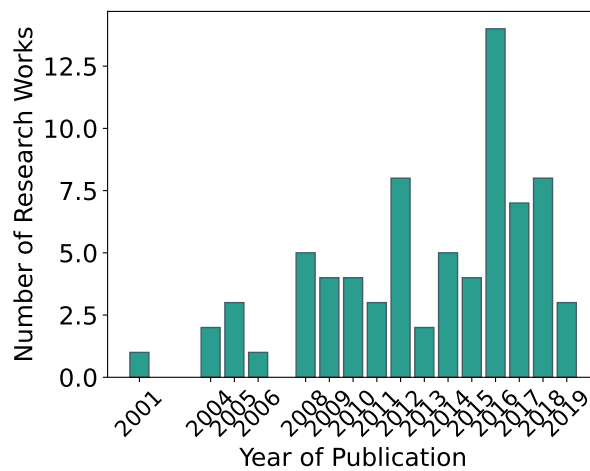


Figure 3.1: Histogram of the number of publications surveyed per year of publication.

3.2.1 State of The Art

Since its inception, the field of emotion Recognition has grown tremendously, with 346,00 results in Google Scholar, and 9,927 in Scopus (searching for "Emotion Recognition"). These numbers illustrate the growing interest in the field and its diversity, requiring a comprehensive review to understand the current state of the art.

Comparing the performance of the research papers is a difficult task since they often differ in the classifiers, the datasets used to train and test the model, sample sizes, the form of validation and the extracted features and signal modalities used. With this in mind, Table 3.1 provides a summary of the state-of-the-art research studies in the field of emotion recognition across its main characteristics.

Table 3.1: Summary of state-of-the-art research studies in the field of emotion recognition and its main characteristics in terms of author, publication year, stimuli used, localization, participant count (#), subject dependency (SD), emotion labels (Labels), employed modalities, choice of classifier, and validation techniques—including Leave-One-Out (LOO), Leave-One-Subject-Out (LOSO), and 10-Fold Cross-Validation (CV), and recognition rates (Rec Rate). Table adapted from [160, 161].

Author	Year	Stim	Loc	#	SD	Labels	Modalities	Classifier	Val	Rec Rate
Petrantonakis P C, et al. [162]	2010	IAPS	L	16	No	Happiness, surprise, anger, fear, disgust, sadness	EEG	KNN, QDA, MD, SVM	LNO	85.17%
Samara A, et al.[163]	2016	DEAP	L	32	Yes	Arousal, valence	EEG	SVM	10-fold	79.83%; 60.43%
Jianhai Zhang et al.[164]	2016	DEAP	L	32	Yes	Arousal, valence	EEG	PNN, SVM	10-fold	PNN: 81.76%; SVM: 82.00%
Ping Gong et al. [165]	2016	Music	L		Yes	Joy, anger, sadness, pleasure Terrible, love, hate, sentimental, lovely, happy,	ECG, EMG, RSP, GSR	DT		92%
Gyanendra K.Verma [166]	2014	DEAP	L	32	Yes	fun, shock, cheerful, Depressing, exciting, melancholy, mellow	EEG, ECG, GSR, EMG, EOG, RESP, SKT, face video	SVM, MLP, KNN, MMC	10-fold	EEG: 81%; peripheral signals: 78%
Vitaliy Kolodyazhniy et al. [167]	2011	Film clips	L	34	Both	Fear, sadness, neutral	ECG, GSR, RSP, SKT, EMG	KNN, MLP, QDA, LDA, RBNF		subj-dep: 81.90%; subj-indep: 78.9%
Dongmin Shin et al. [168]	2017	Videos	L	30	Yes	Amusement, fear, sadness, joy, anger, and disgust	EEG, ECG	BN		98.06%
Foteini Agrafioti et al. [169]	2012	IAPS, video game	L	44	No	Valence, arousal	ECG	LDA	LOO	Arousal: bipartition 76.19%; C.36% valence: 52%-89%
Wanhui Wen et al. [170]	2014	Videos	L	101	No	Amusement, grief, anger, fear, baseline	OXY, GSR, ECG	RF	LOO	74%, LOO

Continued on next page

Table 3.1 continued from previous page

Author	Year	Stim	Loc	#	SD	Labels	Modalities	Classifier	Val	Rec Rate
Jonghwa Kim et al. [171]	2008	Music	F	3	Both	Valence, arousal	ECG, EMG, RSP, SC	pLDA	LOO	Subj-dep: 95%; subj-ind: 77%
Cong Zong et al.[172]	2009	Music (AuBT)	1		Yes	Joy, anger, sadness and pleasure	ECG, EMG, SC, RSP	SVM	10-fold	76%
Valenza et al. [173]	2012	IAPS	L	35	No	Valence, arousal	ECG, EDR, RSP	QDA	40-fold	>90%
Wee Ming Wong et al.[174]	2010	Music (AuBT)	L		Yes	Joy, anger, sadness, pleasure	ECG, EMG, SC, RSP	PSO of synergetic neural classifier (PSO-SNC)	LOO	SBS: 86%; SFS: 86%
Leila Mirmohamadsadeghi et. al. [175]	2016	DEAP	L	32	Yes	Valence, arousal	EMG, RSP	SVM	LOVO	Valence: 74%, arousal: 74%, liking: 76%
Chi-Keng Wu et al.[176]	2012	Film clips	L	33	Yes	Love, sadness, joy, anger, fear	RSP	KNN5	LOO	88%
Xiang Li et al.[177]	2016	DEAP	L	32	Yes	Valence, arousal	EEG	LSTM	5-fold	Valence: 72.06%, arousal: 74.12
Zied Guendil et al. [178]	2015	Music (AuBT)	L		Yes	Joy, anger, sadness, pleasure	EMG, RESP, ECG, SC	SVM		95%
Yuan-Pin Lin et al. [179]	2010	Music	L	26	No	Joy, anger, sadness, pleasure	EEG	MLP, SVM	10-fold	82.29%
Bo Cheng et al. [180]	2008	Music (AuBT)	L		Yes	Joy, anger, sadnes, pleasure	EMG	BP		75%
Saikat Basu et al. [181]	2015	IAPS	L	30	Yes	Valence, arousal (HVHA, HVLA, LVHA, LVLA)	GSR, HR, RESP, SKT	LDA, QDA	LOO	HVHA: 98%, HVLA: 96%, LVHA: 93%, LVLA: 97%
Ingxin Liu et al. [182]	2016	DEAP	L	32	Yes	Valence, arousal	EEG	KNN, RF	10-fold	Valence: 69.9%, arousal: 71.2%

Continued on next page

Table 3.1 continued from previous page

Author	Year	Stim	Loc	#	SD	Labels	Modalities	Classifier	Val	Rec Rate
Mahdis Monajati et al. [183]	2012	Shock test	L	13	Yes	Negative, neutral	GSR, HR, RSP	Fuzzy adaptive resonance theory		94%
Lan Z et al. [184]	2016	IADS	L	5	Yes	Positive, negative	EEG	SVM	5-fold	73.10%
Zheng W L et al. [185]	2018	DEAP + video	L	47	Yes	HAHV HALV LAHV LALV	EEG	G extreme Learning Machine	5-fold	DEAP: 69.67%, SEED: 91.07%
Picard et al. [186]	2001	Clynes protocol	L	1	Yes	Neutral, anger, hate, grief, joy, platonic/romantic love, reverence	EDA, EMG, PPG, RESP	KNN	LOO	81%
Haag et al. [187]	2004	IAPS	L	1	Yes	Low/medium/high arousal and positive/negative valence	ECG, EDA, EMG, SKT, PPG, RESP	NN	3-fold	AR: <96%, VA: <90%
Lisetti and Nasoz[188]	2004	Movie clips and difficult mathematics questions	L	14		Sadness, anger, fear, surprise, frustration, amusement	ECG, EDA, TEMP	kNN; LDA; NN	LOO	72%; 75%; 84%
Healey and Picard [189]	2005	Driving	FC	24		3 stress levels	ECG, EDA, EMG, RESP	LDF	LOO	97%
Leon et al. [190]	2017	IAPS	L	9	No	Neutral/positive/negative valence	EDA, HR, BP	NN	LOSO	71%
Zhai and Barreto [191]	2006	Paced stroop test	L	32		Relaxed and stressed	EDA, PD, PPG, TEMP	NB; DT; SVM	20-fold	79%; 88%; 90%
Kim and André [171]	2008	Music	L	3	Both	HAHV HALV LAHV LALV	ECG, EDA, EMG, RESP	LDA	LOO	subj-dep: 95%, subj-indep: 70%

Continued on next page

Table 3.1 continued from previous page

Author	Year	Stim	Loc	#	SD	Labels	Modalities	Classifier	Val	Rec Rate
Katsis et al. [192]	2008	Simulated driving	L	10		High-low stress, disappointment, euphoria	ECG, EDA, EMG, RESP	SVM; ANFIS	10-fold	79%;77%
Calvo et al. [193]	2009	Clynes protocol	L	3	Both	Neutral, anger, hate, grief, joy, platonic/romantic love, reverence	ECG, EMG	FT; NB; BN; NN; LR, SVM	10-fold	one subject: 37%-98%, all subjects: 23%-71%
Chanel et al. [194]	2009	Recall	L	10		Positively/negatively excited, calm-neutral (in valence-arousal space)	BP, EEG, EDA, PPG, RESP	LDA, QDA, SVM	LOSO	<50%; <47%; <50%, binary: <70%
Khalili and Moradi [195]	2009	IAPS	L	5		Positively/negatively excited, calm (in valence-arousal space)	BP, EEG, EDA, RESP, TEMP	QDA	LOO	66.66%
Healey et al. [196]	2010	Daily-living	F	19		Points in valence arousal space, moods	ACC, EDA, HR, audio	BN; NB; AB; DT	10-fold	
Plarre et al. [197]	2011	Public speaking, mental arithmetic, cold pressor	L/F	21/17		Baseline, different types of stress (social, cognitive, and physical), perceived stress	ACC, ECG, EDA, RESP, TEMP	DT; AB; SVM/HMM	10-fold	82%, 88%, 88%
Hernandez et al. [198]	2011	Calls	F	9	Both	Detect stressful calls	EDA	SVM	LOSO	73%
Valenza et al. [199]	2012	IAPS	L	35	No	5 classes of arousal and five valence levels	ECG, EDA, RESP	QDA	40-fold	>90%

Continued on next page

Table 3.1 continued from previous page

Author	Year	Stim	Loc	#	SD	Labels	Modalities	Classifier	Val	Rec Rate
Koelstra et al. [89]	2012	DEAP	L	32	No	HAHV HALV LAHV LALV	ECG, EDA, EEG, EMG, EOG, RESP, TEMP, facial video	NB	LOSO	AR/VA/LI: 57%/63%/59%
Soleymani et al. [126]	2012	MAHNOB- HCI	L	27	No	Neutral, anxiety, amusement, sadness, joy, disgust, anger, surprise, fear	ECG, EDA, EEG, RESP, TEMP	SVM	LOSO	VA: 46%, AR: 46%
Sano and Picard [200]	2013	Daily-living	F	18	Yes	Stress, neutral	ACC, EDA, phone usage	SVM, kNN	10-fold	<88%
Martinez et al. [201]	2013	Video- game (maze-ball)	L	36	No	Relaxation, anxiety, excitement, fun	EDA, PPG	NN	10-fold	learned features: <75%, hand-crafted: <69%
Valenza et al. [199]	2014	IAPS	L	30	Yes	HAHV HALV LAHV LALV	ECG	SVM	LOO	Valence: 79%, arousal: 84%
Adams et al. [202]	2014	Daily-living	F	7	Yes	Stress, neutral (aroused, non-aroused)	EDA, audio	GMM		74%
Hovsepian et al. [203]	2015	Socioevalu- ative, cognitive, and physical challenges	L/F	26/20	No	Stress, neutral	ECG, resp	SVM/BN	LOSO	92%/>40%
Abadi et al. [130]	2015	DECAF	L	30		High/Low valence, arousal, and dominance	ECG, EOG, EMG, near-infrared face video, MEG	NB, SVM	LOTO	VA/AR/DO: 50-60%

Continued on next page

Table 3.1 continued from previous page

Author	Year	Stim	Loc	#	SD	Labels	Modalities	Classifier	Val	Rec Rate
Rubin et al.[204]	2016	Daily-living	F	10		Panic attack	PA; GB; DT; RR; SVM; RF; kNN; LR	ACC, ECG, RESP	10-fold	bin. panic: 73%-97% bin. pre-panic: 71% - 91%
Jaques et al. [205]	2016	Daily-living	F	30	No	Stress, happiness, health values	SVM; LR; NN;	EDA, TEMP, ACC, phone usage	5-fold	<76%; <86%; <88%
Zenonos et al. [206]	2016	Daily-living	F	4	No	Excited, happy, calm, tired, bored, sad, stressed, angry	ACC, ECG, PPG, SKT	kNN, DT, RF	LOSO	58%; 57%; 62%
Gjoreski et al. [207]	2017	Daily-living	L/F	21/ 5	No	Lab: no/low/high stress; field: stress, neutral	ACC, GSR, BVP, SKT	SVM, RF, AB, kNN, BN, DT	LOSO	<73%/ <90%
Mozos et al. [208]	2017	TSST	L	18		Stress, neutral	ACC, GSR, BVP, audio	AB, SVM, kNN	CV	94%; 93%; 87%
Schmidt et al. [88]	2018	WESAD	L	15	No	Neutral, fun, stress	ACC, ECG, GSR, EMG, RESP, SKT, BVP	DT, RF, kNN, LDA, AB	LOSO	<80%/<93%
Hao Tang et al.[209]	2017	DEAP	L	32		Arousal, valence	EEG, ECG, GSR, EMG, EOG, RESP, SKT, face video	Bimodal-LSTM	10-fold	Arousal: 83.23%, valence:83.83%
Wei Liu et al.[210]	2016	DEAP	L	32		Positive, neutral, negative	EEG	BDAE		83.25%
Tripathi et al. [211]	2017	DEAP	L	32		Valence, arousal	EEG	DNN, CNN		CNN: (V)81.406%, 73.36%(A); (DNN) valence: 75.78%, arousal: 73.125%
Wenqian Lin et al.[212]	2017	DEAP	L	32	Yes	Valence, arousal	EEG, ECG, GSR, EMG, EOG, RESP, SKT, face video	CNN	10-fold	Arousal: 87.30%, valence: 85.50%

Continued on next page

Table 3.1 continued from previous page

Author	Year	Stim	Loc	#	SD	Labels	Modalities	Classifier	Val	Rec Rate
Santamaria-Granados et al. [213]	2019	AMIGOS	L	40		Valence, arousal	EEG + ECG	DCNN		Arousal: 0.76, valence: 0.75
Subramanian et al. [127]	2018	ASCERTAIN	L	58		Arousal, Valence	EEG, ECG, GSR, facial activity data	SVM, NB	LOO	(GSR,NB) Arousal: 0.68, valence: 0.68
Lee et al. [214]	2018	Movie	L	50	Yes	Negative, neutral emotions	ECG, SKT, EDA	NN, LDA, QDA	LOO	NN: 92.5%
Yang et al. [215]	2018	Video game	L	58		Arousal, valence	ECG, EDA, RESP, EMG, ACC	SVM, RBF SVM, DT, RF	10-fold	Arousal: 0.559, valence: 0.524
Li et al. [216]	2019	DEAP, stroop test	L	32 + 20	Both	Low, medium and high stress	BVP, GSR	LR, eSVR, CNN, ST-SVR	CV	F1-score between 0.943, 0.970 and 0.984
Zhao et al. [212]	2018	ASCERTAIN	L	58	No	Arousal, valence	GSR, EEG, ECG, facial landmarks	Vertex-weighted Multi-modal Multi-task Hypergraph Learning, SVM, NB, hypergraph	10-fold	(VM2HL) Valence: 74.34, arousal: 79.46
Anusha et al. [217]	2018	TSST, Stroop Color Word test, Mental Arithmetic test	L	34	No	Baseline, stress	EDA, ECG, SKT	LDA, QDA, SVM, 3-NN,	LOSO	(EDA+SKT) 97.13%
Sirisha Devi et al. [218]	2019	DEAP	L	50		Valence, arousal	EEG, HR, GSR, RESP	LDA		93.8%
Xia et al. [219]	2018	Stress	L	22		Stress, control	EEG, ECG	SVM-sigmoid, SVM-RF	10-fold	79.54%

Continued on next page

Table 3.1 continued from previous page

Author	Year	Stim	Loc	#	SD	Labels	Modalities	Classifier	Val	Rec Rate
Agrafioti et al. [169]	2012	IAPS	L	31	Yes	Neutral, gore, fear, disgust, excitement, erotica, game elicited mental arousal	ECG	LDA	LOO	active/pas AR: 78/52% positive/neg VA: <62%
Han-Wen Guo et al. [220]	2016	Movie clips	L	25	Yes	Positive, negative	ECG	SVM		71.40%
Hernan F. Garcia et al. [221]	2016	DEAP	L	32	Yes	Valence, arousal	EEG, EMG, EOG, GSR, RSP, T, BVP	SVM		Valence: 88.33%, arousal: 90.56%
Liu et al. [222]	2005	Cognitive tasks.	L	15		Anxiety, boredom, engagement, frustration, anger	ECG, EDA, EMG	kNN; RT; BN; SVM	LOO	75%; 84%; 74%; 85%
Wagner et al. [223]	2005	Music (AuBT)	L	1	Yes	Joy, anger, pleasure, sadness	ECG, EDA, EMG, RESP	kNN; LDF; NN	LOO	81%; 80%; 81%
Zhu et al. [224]	2016	Daily-living	F	18	No	Angle in valence arousal space	ACC, phone context	RR	LOSO	
Birjandtalab et al. [225]	2016	Physical, cognitive, emotional stress	L	20		Relaxation, physical, emotional, cognitive stress	ACC, EDA, TEMP, HR, SpO2		GMM	<85%

Upon the survey, the following observations can be made:

- **Elicitation Material:** Figure 3.2a shows a histogram with the number of publications surveyed per elicitation material. As can be seen, videos and films are the most commonly used elicitation material.

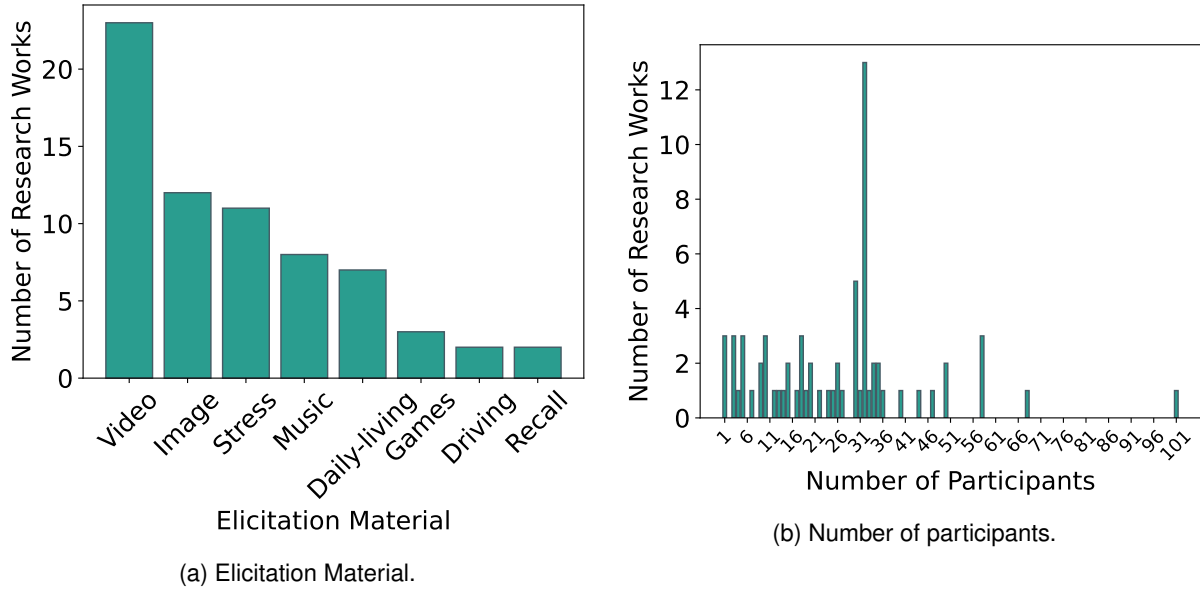


Figure 3.2: Histogram of the number of publications surveyed for this document per elicitation material and number of participants used in the datasets. Figures extracted from [9].

- **Constrained vs. Unconstrained setting:** Most studies are performed in a lab setting, and these, on average, achieve higher accuracy. Lab studies are devised to elicit specific emotions, pre-validated and easily acquired and replicated in an elevated number of subjects with quality ground-truth annotation. Additionally, often the subjects are asked to remain still, thus, minimising movement artefacts.
- **Number of Subjects:** Figure 3.2b shows a histogram with the number of publications surveyed per number of subjects. As can be seen, the majority of the surveyed publications reported data for between 1 to 50 participants.
- **Subject-dependent vs. Subject-independent:** Subject-dependent algorithms achieve on average higher results than subject-independent. This can be explained by physiological data and emotions being biased by the subject's physiological internal and external factors.
- **Emotion Models:** Most works focus on the implementation of binary classification techniques, separating arousal from valence and stress from no-stress activities.
- **Signal Modalities:** Figure 3.3a shows a histogram with the number of publications surveyed per signal modality. As can be seen, EDA is the most commonly used signal for emotion recognition,

followed by ECG, respiration, and EEG. Most research studies reported that the classification performance increases with the increment of the number of signal modalities.

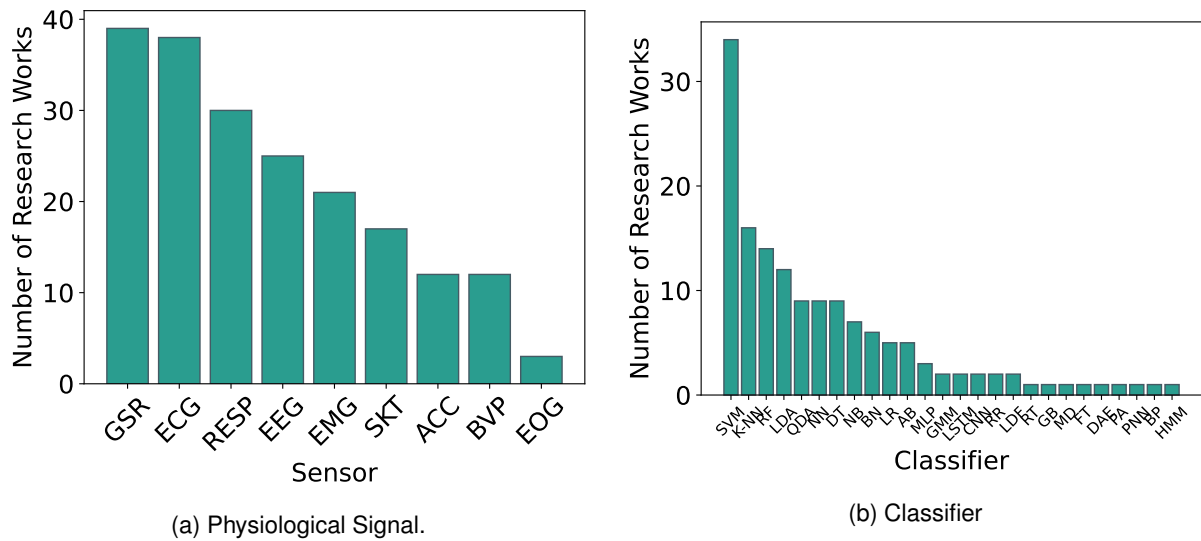


Figure 3.3: HHistogram with the number of publications across physiological signals and classifiers.

- Classifiers:** Figure 3.3b shows a histogram with the number of publications surveyed per classifier. As can be seen, the Support Vector Machine (SVM) is the most prevalent classifier, followed by K-Nearest Neighbors (K-NN), Random Forest, and Linear discriminant analysis. The SVM algorithm is the most prevalent algorithm in the literature, showing good results and low computational complexity. Non-traditional deep learning techniques mostly use EEG data, which can be related to more data being available.
- Dimensionality Reduction:** The application of feature selection and data dimensionality reduction algorithms is reported across the literature [167, 169, 181]. These methods allow to increase the classification performance, however, at the expense of increased time and computational cost.
- Validation Techniques:** To avoid overfitting many works apply K-fold CV, leave one sample out, or LOSO techniques. The latter allows to obtain subject-independent evaluations, leading to more generalised results.
- Evaluation Metrics:** Accuracy is the most commonly applied metric to evaluate the model's performance. More detailed metrics, that consider data imbalance and allow the analysis of each class label separately such as F1-score, Precision and Recall are also reported in the literature
- Class Labels:** Most classification tasks are performed on the arousal and valence dimensions. The stress dimension is also explored in the literature.

Over the past years, advancements in the fields of emotion psychology, computer science and electronics have endured the growth of affective computing theory and research. These interdisciplinary

efforts have provided a deeper understanding of emotions, and the development of ubiquitous, fast and pervasive wearable technology. This culminated in the development of new datasets and the development of more intricate machine-learning algorithms, marking a notable step forward in the field.

3.2.2 Machine Learning

The review of the current literature on affective computing showed that most works resort to machine learning algorithms to perform emotion classification [9]. Machine learning algorithms can be divided into supervised and unsupervised. In supervised learning [226], a model (h) learns to mathematically characterize an annotated training set, with the sample (x) and its label (y), with the aim that given a new unlabelled sample (x_{new}), the model can predict its class ($\hat{y} = h(x_{new})$). The sample is described through a feature vector, which mathematically characterizes the sample in a multidimensional space. Each element in the vector represents a particular feature (or attribute) of the sample.

Moreover, the field is divided into traditional machine learning and deep learning approaches. Deep learning algorithms perform repeated weighted combinations of the input features followed by a non-linear transformation. Common deep learning algorithms are feedforward neural networks, Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), autoencoders, and transformers, among others. Deep learning has been shown to achieve state-of-the-art results in many fields, including emotion recognition. However, they require large datasets to attain accurate predictions and are more complex to train.

As the previous survey showed, traditional machine learning algorithms are still widely used in the field of emotion recognition. The most prevalent algorithms are the SVM, K-NN, Random Forest, and Linear discriminant analysis. The classification performance is dependent on the physiological signal used as input to learn the model, its labels, and the dataset data. Across the literature, EDA was the most prevalent signal for emotion recognition, closely followed by ECG, respiration, and EEG.

Moreover, the literature denoted that higher accuracy can be achieved by the exploration of multi-modal ANS data [70, 9]. The use of simultaneous physiological data is done through two main methods [10]: feature fusion [227] and decision fusion [89, 228]. In feature fusion, a feature vector is extracted from each modality and is concatenated to form one main feature vector. The concatenated feature vector is then given as input to the machine learning algorithm. In decision fusion, a decision/classification is performed for each modality. Then, the outputs of each classifier are combined to calculate a final prediction by a voting system. Common voting systems are majority voting, or weighted majority voting [10]. Both techniques are observed in the state of the art, but there is no clear indication of which method will return the most accurate prediction.

The current section focuses on the use of traditional machine learning models for emotion recognition to benchmark the current state-of-the-art datasets, signal modalities, and classification. Multimodal

data fusion techniques will also be explored to understand how to better deal with multiple sources of physiological data.

3.3 Methods

To perform the emotion classification task, a data processing pipeline was defined (Figure 3.4). The pipeline is divided into five steps: data loading, signal pre-processing; feature extraction and selection; classification; and evaluation. Each step is detailed in the following subsections.

3.3.1 Data

Five publicly available datasets for emotion recognition have been identified, namely: ITMDER [135]; WESAD [88]; DEAP [89]; MAHNOB [126]; and EESD [128]. The datasets were introduced in Section 2.3 and are further detailed in Table 3.2.

Table 3.2: Public datasets explored for emotion classification. The datasets are described by their: classes; percentage of the number (#) of samples per class (binary ground-truth class 0 and 1), for Arousal (A) and Valence (V); Demographic Information (DI) with the number of participants, their age (in years), and gender (Female (F), Male (M)); Device; Sampling rate.

Dataset	Classes	# Samples per Class	DI	Device	Sampling Rate
ITMDER [135]	Low-high arousal/valence	A: 0.54 (0); 0.46 (1) V: 0.12 (0); 0.88 (1)	18 23 ± 3.7 10 (F) – 13 (M)	Chest strap and armband based on BI Talino [229]	1000
WESAD [88]	Neutral, Stress, Amusement + 4 Questionnaires	A: 0.86 (0); 0.14 (1) V: 0.07 (0); 0.93 (1)	15 27.5 ± 2.4 3 (F) – 12 (M)	RespiBAN [229], Empatica E4 ¹	ECG, RESP: 700; EDA: 4; PPG: 64
DEAP [89]	arousal, valence, Like/dislike, Dominance and Familiarity	A: 0.41 (0); 0.59 (1) V: 0.43 (0); 0.57 (1)	32 16 (F) – 16 (M) 19 – 37	Biosemi Active II ²	128
MAHNOB [126]	arousal, valence, Dominance	A: 0.48 (0); 0.52 (1) V: 0.47 (0); 0.53 (1)	27 26.06 ± 4.39 17(F) – 13(M)	Biosemi Active II system	256
Eight-Emotion (EESD) [128]	Neutral, Anger, Hate, Grief, Platonic love, Romantic Love, Joy, and Reverence	A: 0.5 (0); 0.5 (1) V: 0.5 (0); 0.5 (1)	1 1 (F)	Thought Technologies ProComp ³	256

¹ <https://www.empatica.com/research/e4/>; Accessed on 31/03/2024

² <https://www.biosemi.com>; Accessed on 20/02/2024

³ <https://thoughttechnology.com/procomp-infiniti-system-w-biograph-infiniti-software-t7500m/>; Accessed on 20/02/2024

The emotion recognition problem was defined as two binary classification tasks, for the valence and arousal dimensions, with class 0 for a negative emotion and class 1 for a positive emotion. The class labels were translated from the original representation to the 0, 1 classes taking the neutral as 0.

Moreover, taking the arousal and valence dimensions. It can be noticed that the datasets are heavily imbalanced, namely the WESAD valence dimension (7% (class 0); 93% (class 1)).

3.3.2 Signal Pre-Processing

Physiological signals usually contain noise and artefacts that can affect the performance of the emotion recognition system. With this in mind, a pre-processing step was applied to remove noise and artefacts. Common pre-processing steps include filtering and normalization. Filtering is used to remove noise from the signal. The following filters were applied:

- **Electrocardiography:** Finite Impulse Response (FIR) band-pass filter of 300^{th} order and 3 — 45 Hz cut-off frequency.
- **Electrodermal Activity:** Butterworth low-pass pass filter of 4^{th} order and 1 Hz cut-off frequency.
- **Respiration:** Butterworth band-pass filter of 2^{nd} order and 0.1 — 0.35 Hz cut-off frequency.
- **Photoplethysmography:** Butterworth band-pass filter of 4^{th} order and 1 — 8 Hz cut-off frequency.

After noise removal, to reduce the inter-subject variability, the data was normalized per user by $\frac{X-\mu}{\sigma}$; μ : sample average, σ : sample standard deviation; X : physiological data, to $X \in [0, 1]$ per subject, Lastly, the data was segmented into 40-second sliding windows with 75% overlap.

3.3.3 Feature Extraction and Selection

After the data cleaning step, features are extracted from the signals. These follow characteristic signal properties, e.g. rise time in EDA or HR in ECG, as well as domain-related features, such as time-domain; frequency-domain; statistical or non-linear features. Time-domain features are extracted from the signal in the time domain. Frequency-domain features are extracted from the signal in the frequency domain. Non-linear features are extracted from the signal in the time domain, but they are not linearly related to the signal.

In total, 570 EDA-based, 373 PPG-based, 322 ECG-based, and 487 respiration-based features were extracted from each sensor, respectively. The features are described in



Figure 3.4: Typical emotion recognition data classification processing pipeline.

3.3.4 Classification

From the total of set of extracted features, two methodologies were implemented: feature fusion and decision fusion. In feature fusion, the data from the diverse sensors was concatenated to form a single feature vector. In decision fusion, an independent classification is performed for each modality and the final prediction is obtained by a weighted majority voting system. The weights were obtained by the classifier accuracy on the validation set (described below), giving more weight to the more accurate classifiers.

Traditional supervised classifiers were tested, namely: K-NN; Decision Tree; Random Forest; SVM; AdaBoost; Gaussian Naive Bayes; and Quadratic Discriminant Analysis. More details about these classifiers can be found in [230] and references therein.

3.3.5 Evaluation

The hyper-parameters for each classifier were optimized using a 4-fold CV on the validation set (Table 3.3).

Table 3.3: Hyper-parameters and respective search space across classifiers. The hyper-parameters were optimized using a grid search with 4-fold CV on the validation set. Nomenclature: number of features (# features). Information on the parameters can be found in the scikit-learn library [231].

Classifier	Parameter	Value
Decision Tree	criterion	{"gini", "entropy"}
	splitter	{"best", "random"}
	min_samples_split	[2, 20]
	min_samples_leaf	[1, 20]
	max_depth	[1, 20]
Nearest Neighbors	n_neighbors	[1, 10]
	leaf_size	[1, 10]
	p	[1, 10]
Random Forest	max_depth	[1, # features]
	max_features	[1, # features]
	min_samples_split	[2, # features]
	min_samples_leaf	[2, # features]
	bootstrap	[True, False]
	criterion	{"gini", "entropy"}
	n_estimators	[5, 20]

The 4-fold CV was selected to optimize the number of iterations and the homogeneity in the number of classes in the training and validation sets. The final classifier evaluation was performed using LOSO. The classifier with the best performance was used to test and benchmark the emotion recognition performance using decision and feature fusion. The performance of each method was measured using: Accuracy — $\frac{TP+TN}{TP+TN+FP+FN}$; Precision — $\frac{TP}{TP+FP}$; Recall — TP; F1-score — the harmonic mean of precision and recall, with default average set to "binary". Nomenclature: TP — True-Positive; FP — False-Positive; FN — False-Negative.

3.4 Results

Results were divided to describe first the performance for each single modality and afterwards the multi-modality approach.

3.4.1 Single Modality Models

Table 3.9 shows the results for each sensor modality individually. The results not covered by the literature were left as empty cells, i.e. in the cases where a dataset did not have the data for a physiological sensor. The EDA is divided into data collected on the fingers (*EDA F row*) and in the hand (*EDA H row*).

Starting the analysis with the ITMDER dataset, it can be seen that data was available for all the sensors and, throughout the different modalities, the proposed methodology was able to surpass the state-of-the-art (*SoA* column) or obtain very competitive results for both the arousal and valence dimensions. For the arousal dimension, the method trained on ECG data (in bold) outperformed the remaining physiological signals. Across all sensors, the valence dimension attains the highest prediction scores, but no modality stands out.

For the WESAD dataset, in the arousal dimension, the F1-score drops to 0.0 compared to the accuracy value. This means that the class labels were largely imbalanced with none of the test sets having one of the labels. Analysing all the modalities, the PPG obtained the best performance for the valence dimension.

For the remaining datasets, no results were reported with comparable configurations to the ones performed in the current work, hence no comparison is shown.

- **Classifiers:** The classifiers selected by the grid search are shown in Table 3.4. Across datasets, different classifiers were selected by their performance. In general, Random Forest and Adaboost stand out being selected 12 and 11 times, respectively.

Table 3.4: Single-modality experimental results best performing supervised learning classifier per dataset, sensor modality and emotion dimension. Evaluation obtained using 4-fold CV. Nomenclature: K-NN; Decision Tree (DT); Random Forest (RF); SVM; Gaussian Naive Bayes (GNB); and Quadratic Discriminant Analysis (QDA).

	ITMDER		WESAD		DEAP		MAHNOB-HCI		EESD	
	Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence
EDA H	DT	RF	RF	RF	SVM	SVM	AdaBoost	SVM	AdaBoost	AdaBoost
EDA F	AdaBoost	QDA								
ECG	AdaBoost	RF	QDA	RF			RF	AdaBoost		
PPG	QDA	RF	AdaBoost	RF	RF	RF			AdaBoost	AdaBoost
Resp	AdaBoost	RF	RF	RF	AdaBoost	RF	QDA	AdaBoost	AdaBoost	QDA

- **Features:** The features obtained by the forward feature selection algorithm are shown in Table 3.5 and Table 3.6 for the arousal and valence dimensions, respectively. As it can be seen, similar

features are selected across the different sensor modalities, exploring correlated aspects of each modality, possibly explaining why no algorithm stands out in accuracy.

Figure 1 shows the features selected for the supervised learning algorithm (Table 3.4 with its corresponding results shown in Table 3.9). As can be seen, most of the features are selected once per dataset (value of 1 in the histogram). The exceptions are the features EDA onsets spectrum mean and PPG signal average which are selected twice for the arousal dimension. For the valence dimension, the features EDA onsets spectrum mean value are selected 4 times, respiration signal means 3 times, PPG signal mean 2 times, and ECG NNI (NN intervals) minimum peaks value twice.

3.4.2 Multi-Modality Models

The experimental results for the decision fusion and feature fusion algorithms are shown in Table 3.10.

Starting with the ITMDER dataset, both the decision fusion and feature fusion methods outperform the literature (*SoA* column). Similarly to the single modality classification, higher accuracy is obtained for the valence dimension.

For the DEAP dataset, both techniques attain a similar performance and surpass the state of the art for the arousal dimension and valence F1-score.

For the MAHNOB dataset, the proposed methodology is not able to surpass the results of the literature, however, competitive results are obtained.

In the EESD dataset, decision fusion obtains a better performance compared to feature fusion. However, the performance is not coherent for the remaining datasets. For the EESD and WESAD no comparable results were identified in the literature in equivalent dimensions.

Regarding the computation time (*T row*), feature fusion is faster than decision fusion and, overall, shows lower computational complexity. The results highlight an average execution time of up to two orders of magnitude lower compared to decision fusion. The computation was performed using Python 3.7.4; in a computer with Memory: 16 GB 2133 MHz LPDDR3 and Processor of 2.9 GHz Intel Core i7 quadruple core. This factor along with the similar prediction accuracy across the two methods (with no multi-modality method standing out) highlights the feature fusion method which will be further analyzed next.

Table 3.7 shows the classifiers used in each dataset per dimension for the feature fusion method. Once again, the results are heavily dependent on the algorithm, with no classifier standing out and being selected more than the average. Table 3.8 shows the feature selected in each classifier (shown in Table 3.7 with the results shown in Table 3.10). The experimental results show that, similarly to supervised learning, most features are fitted for each dataset, with none in common being selected across the datasets.

Table 3.5: Features selected for the arousal dimension in the single-modality methods, i.e. supervised learning and decision fusion. Results obtained using 4-fold CV.

	ITMDER	WESAD	DEAP	MAHNOB-HCI	EESD
EDA H	peaksOnVol_minpeaks EDRVoIRatio_iqr onsets_temp_dev	EDA_onsets_spectrum_mean	onsets_spectrum_mean	half_rec_minAmp half_rec_rms amplitude_dist onsets_spectrum_statistic_hist43 rise_ts_temp_curve_distance phasic_rate_maxpeaks onsets_spectrum_meddiff EDRVoIRatio_zero_cross	phasic_rate_abs_dev onsetspeaksVol_minpeaks
EDA F	onsets_spectrum_statistic_hist81 peaksOnVol_iqr six_rise_autocorr				
ECG	statistic_hist73, statistic_hist115 hr_sadiff statistic_hist7 statistic_hist137	mean rpeaks_medadev hr_meandiff		hr_mindiff	
PPG	hr_max hr_meandiff	mean	mean		spectral_skewness temp_curve_distance statistic_hist18 statistic_hist13 statistic_hist15 meddiff
RESP	exhale_counter inhExhRatio_iqr	statistic_hist0	mean	hr_total_energy meandiff statistic_hist95 inhale_dur_temp_curve_distance statistic_hist27 hr_meandiff	exhale_meanadiff max, zeros_mean

Table 3.6: Features selected for the valence dimension in the single-modality methods, i.e. supervised learning and decision fusion. Results obtained using 4-fold CV.

	ITMDER	WESAD	DEAP	MAHNOB-HCI	EESD
EDA H	onsets_spectrum_mean rise_ts_temp_curve_distance rise_ts_medadev	onsets_spectrum_mean	onsets_spectrum_mean	onsets_spectrum_mean	amplitude_mean onsets_spectrum_meanadev half_rise_medadev onsets_spectrum_statistic_hist9 EDRVolRatio_medadiff half_rec_minpeaks
EDA F	onset_peaks_Vol_max half_rise_mean, peaks_max onsets_spectrum_statistic_hist120 half_rec_meandiff onsets_spectrum_statistic_hist91 half_rise_var peaks_Onset_Vol_skewness				
ECG	nni_minpeaks	nni_minpeaks statistic_hist95		rpeaks_meandiff max mindiff	
PPG	statistic_hist44 meanadiff hr_meanadiff onsets_mean hr_meandiff	median minAmp	mean		mean statistic_hist16 statistic_hist5 statistic_hist31 meddiff
Resp	mean exhale_median statistic_hist196	mean	mean	hr_maxpeaks statistic_hist55 zeros_skewness statistic_hist36	iqr

Table 3.7: Classifier used in the multi-modal feature fusion method for the arousal and valence dimension. Results obtained using 4-fold CV.

Classifier	ITMDER		WESAD		DEAP		MAHNOB-HCI		EESD	
	Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence
	SVM	RF	QDA	SVM	QDA	GNB	GNB	QDA	DT	RF

Table 3.8: Features used in the multi-modal feature fusion method for the arousal and valence dimension. Results obtained using 4-fold CV.

ITMDER	WESAD	DEAP	MAHNOB-HCI	EESD
Arousal				
EDA_H_onsets_spectrum_mean	BVP_median ECG_min Resp_statistic_hist64	Resp_zeros_sadiff BVP_statistic_hist29 EDA_phasic_rate_total_energy EDA_rise_ts_mindiff Resp_statistic_hist25	Resp_inhExhRatio_maxpeaks EDA_phasic_rate_iqr Resp_inhExhRatio_zero_cross Resp_inhExhRatio_skewness ECG_rpeaks_meanadiff ECG_minpeaks Resp_meandiff EDA_onsets_spectrum_minAmp EDA_onsets_spectrum_statistic_hist22 ECG_hr_dist EDA_onsets_spectrum_statistic_hist2	Resp_exhale_max EDA_amplitude_kurtosis
Valence				
EDA_H_peaksOnVol_minAmp BVP_mean EDA_F_EDRVolRatio_total_energy EDA_H_onsets_spectrum_statistic_hist112	BVP_median ECG_dist ECG_zero_cross ECG_statistic_hist143	Resp_statistic_hist60 EDA_half_rise_dist BVP_statistic_hist10 BVP_statistic_hist39 EDA_half_rise_temp_curve_distance BVP_hr_maxAmp	ECG_meanadiff EDA_rise_ts_meandiff Resp_inhale_dur_dist EDA_onsets_spectrum_statistic_hist5	EDA_amplitude_mean BVP_statistic_hist35 Resp_rms Resp_zeros_meandiff EDA_onsets_spectrum_statistic_hist22

3.5 Discussion

In response to RQ 1.1, (on exploring the literature sensor modalities, supervised learning classifier, and specific features for emotion recognition) a comprehensive analysis of the literature on emotion recognition using physiological data was performed. The survey identified the most prevalent physiological signals used in the literature as EDA, ECG, and respiration, explored using traditional machine learning models such as the SVM, K-NN, and random forest. The use of this data was explored on predominant datasets in the literature, namely: ITMDER, WESAD, DEAP, MAHNOB, and EESD to perform binary arousal and valence classification. The experimental results showed that very similar results were obtained across classifiers and sensors, possibly due to the redundant features being selected across sensor modalities (see Table 3.5). Further work could focus on exploring different feature extraction and selection methods, or the use of deep learning techniques that can overcome many of the concerns regarding feature extraction and selection.

In response to RQ 1.2, (on what is the expected performance range for emotion classification), the results showed that the performance of emotion recognition is highly dependent on the dataset. The F1-score ranges from a very poor performance score of 0% to an elevated performance of around 97%, depending on the dataset. On the whole, higher prediction scores are obtained for the valence dimension. This could result from the models being fitted to each dataset, with its specific characteristics, sensor types, body sites, and data collection protocols. Future work could also explore transfer learning or the development of a methodology that could be applied across the different dataset setups.

Lastly, regarding RQ 1.3 (on the exploration of the optimal method to deal with multi-modality), the use of feature fusion was explored against decision fusion. Comparing the results of using a single modality to multi-modality, the latter can either preserve or surpass the results from using a single modality, depending on the dataset and dimension, however, the feature fusion was faster and showed a lower computational complexity. These results are in line with the results found in the literature [89],

3.6 Conclusion

The field of affective computing has experienced enormous growth in the latter years [9], however, many of its applications are still in their infancy as the current effort in the field is on the development of emotion recognition algorithms and their underlying structure. This chapter addressed *Obj 1. on the Development of Affective Computing Algorithms*, namely RQ 1.1 to 1.3 which focused on critical aspects to deploy emotion recognition systems, namely the selection of sensor modalities, supervised learning classifiers, and features for emotion recognition, the expected performance range for emotion classification, and the optimal method to deal with multi-modality.

Through this analysis, this part of the thesis benchmarks emotion recognition in terms of low/high arousal and valence through the most prevalent physiological signals identified in the literature [9]: ECG, EDA, respiration and PPG. The experimental results obtained using the proposed methodology show comparable to superior results to those described in the literature. Although very similar results were obtained across the modality fusion algorithms, feature fusion showed a lower processing time (up to two orders of magnitude lower) when compared to decision fusion. Moreover, employing multi-modality by feature fusion either maintained or exceeded the performance of single-modality approaches in emotion recognition.

On the whole, this chapter extends the state-of-the-art by: 1) Benchmarking the emotion recognition classification performance for supervised learning classifiers, modality signals and extracted features in terms of accuracy and F1-score, surpassing the literature for some of the identified datasets; and 2) Analysing multi-modal approaches, namely feature fusion and decision fusion.

Lastly, the analysis revealed a significant limitation in the current state of the art: the public datasets discussed were all collected in laboratory settings and exclusively in individual contexts. This approach overlooks the performance of emotion recognition in real-world scenarios and neglects the influence of group dynamics on emotion recognition. The subsequent chapter (Chapter 4) will shift the focus from individual-level to group-level emotion recognition, aiming to assess the performance of emotion classification within group scenarios.

Table 3.9: Emotion classification single-modality experimental results in terms of the classifier’s accuracy (1st row) and F1-score (2nd row) in %. Results are obtained using LOSO. Nomenclature: State-of-the-art results (SoA, [135]); EDA obtained on a device placed on the hand and finger, respectively (EDA H, EDA F). The best results are shown in bold.

	ITMDER				WESAD				DEAP		MAHNOB-HCI		EESD	
	Arousal	SoA	Valence	SoA	Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence
EDA	59.65 ± 13.46	57.2	89.26 ± 17.3	72.1	85.78 ± 16.55	92.86 ± 11.96	58.91 ± 15.21	56.56 ± 9.07	50.61 ± 21.84	56.43 ± 34.84	59.38 ± 16.24	68.75 ± 18.75		
H	40.74 ± 26.0		93.2 ± 12.37		0.0 ± 0.0	95.86 ± 6.99	72.91 ± 12.92	71.83 ± 7.42	47.53 ± 31.47	64.63 ± 34.57	56.82 ± 20.8	66.71 ± 23.1		
EDA	56.03 ± 11.0	57.2	90.91 ± 11.29	72.1										
F	45.67 ± 20.01		91.24 ± 18.75											
ECG	68.33 ± 5.58	65.6	89.26 ± 17.3	70.0	85.75 ± 16.61	92.86 ± 11.96			49.36 ± 37.5	59.15 ± 24.5				
	58.79 ± 21.54		93.2 ± 12.37		0.0 ± 0.0	95.86 ± 6.99			53.0 ± 39.62	56.58 ± 32.61				
PPG	58.44 ± 12.69	66.0	89.35 ± 17.23	69.5	85.78 ± 16.55	94.39 ± 9.98	58.88 ± 15.19	56.56 ± 9.07			67.5 ± 13.35	66.25 ± 16.35		
	45.91 ± 25.24		93.25 ± 12.34		0.0 ± 0.0	96.68 ± 6.01	72.9 ± 12.91	71.83 ± 7.42			66.98 ± 15.95	64.49 ± 22.07		
RESP	62.37 ± 16.83	58.5	89.26 ± 17.3	62.9	85.78 ± 16.55	92.86 ± 11.96	58.83 ± 14.78	56.56 ± 9.07	50.62 ± 21.25	46.57 ± 20.67	72.5 ± 12.87	67.5 ± 10.0		
	51.79 ± 23.16		93.2 ± 12.37		0.0 ± 0.0	95.86 ± 6.99	72.6 ± 12.74	71.83 ± 7.42	44.28 ± 31.66	48.27 ± 28.44	70.12 ± 15.72	57.92 ± 15.12		

65

Table 3.10: Experimental results for the feature fusion and decision fusion methodologies in terms of Accuracy (A) and F1-score (F1), and time (T) in seconds, per dataset for the arousal dimension. Results obtained using LOSO. The SoA column contains the results found in the literature (ITMDER [135], DEAP [89], MAHNOB-HCI [126]). The best results are shown in bold.

DF	Arousal	ITMDER		SoA	WESAD		Arousal	DEAP		SoA	Arousal	MAHNOB-HCI		SoA	EESD	
		SoA	Valence		Arousal	Valence		SoA	Valence			SoA	Valence		Arousal	Valence
A	66.7 ± 9.0	58.1	89.3 ± 17.3	57.12	85.8 ± 16.5	92.9 ± 12.0	58.9 ± 15.2	56.6 ± 9.1	54.7 ± 13.3		58.1 ± 6.1				75.0 ± 14.8	75.6 ± 17.9
F1	50.9 ± 23.5		93.2 ± 12.4		0.0 ± 0.0	95.9 ± 7.0	72.9 ± 12.9	71.8 ± 7.4	63.8 ± 15.8		68.1 ± 8.9				73.4 ± 16.4	72.4 ± 22.5
T	1.5 ± 0.0		1.35 ± 0.0		2.04 ± 0.0	2.0 ± 0.0	1.58 ± 0.0	1.73 ± 0.0	1.1 ± 0.0		1.35 ± 0.0				0.6 ± 0.0	0.7 ± 0.0
FF																
A	87.6 ± 16.7		89.26 ± 17.3		87.6 ± 16.7	92.9 ± 12.0	60.0 ± 13.9	57.0	56.9 ± 8.2	62.7	55.2 ± 15.4	64.2	68.7	62.7 ± 3.9	60.0 ± 18.4	68.7 ± 22.2
												57	57.5	56.0 ± 10.2		
F1	19.4 ± 34.4		93.2 ± 12.4		19.4 ± 34.4	95.9 ± 7.0	67.3 ± 23.8	53.3	70.7 ± 7.6	60.8	67.5 ± 16.6			59.0 ± 15.1	56.7 ± 22.5	67.7 ± 24.7
T	0.02 ± 0.0		0.02 ± 0.0		0.02 ± 0.0	0.07 ± 0.01	0.02 ± 0.01		0.02 ± 0.0		0.01 ± 0.0			0.01 ± 0.0	0.0 ± 0.0	0.01 ± 0.0

Chapter 4

Group Emotion Recognition

As detailed in Section 2.4, group dynamics such as emotion contagion and mimicry can lead to changes in individual-level emotions and the emergence of macro-level phenomena, the Collective Emotions. These phenomena, akin to individual emotions, can be detected through physiological signals.

The current chapter further contributes to *Obj 1. on the Development of Affective Computing Algorithms*, namely RQ 1.4 and 1.5, by exploring the performance of emotion recognition in group settings. This is achieved by analysing various physiological synchrony measures to measure emotional transfers within groups and proposing a novel approach that integrates the group context into emotion recognition using physiological signals.

The contents within this chapter were adapted, with permission, from:

- P. Bota, T. Zhang, A. El Ali, A. Fred, H. P. da Silva, and P. Cesar, “Group synchrony for emotion recognition using physiological signals,” *IEEE Trans. on Affective Computing*, vol. 14, no. 4, pp. 2614–2625, 2023

Contents

4.1	Introduction	68
4.2	Background	69
4.2.1	Group Emotion Classification	69
4.2.2	Metrics for Physiological Synchrony	70
4.2.3	Physiological Datasets for Group Emotion Recognition	71
4.3	Methods	72
4.3.1	Workflow	73
4.3.2	Synchronisation Metrics	74
4.3.3	Data Representation	75
4.3.4	Classification Models	77

4.4	Results	80
4.4.1	Intrapersonal Model	80
4.4.2	Interpersonal Model	83
4.5	Discussion	85
4.5.1	Synchronisation Metrics and Data Representations	85
4.5.2	Interpersonal Model vs Intrapersonal Model	86
4.6	Conclusion	87

4.1 Introduction

Humans are social beings, spending a large amount of time in collective activities, either at work, for leisure or at home [232]. In such contexts, our emotions are adapted to the group and its members [149, 233]. The use of group information towards emotion classification in audiovisual sources is vastly explored, largely motivated by challenges such as the "Emotion Recognition in the Wild" (EmotiW), which focused on group emotion analysis using images (EmotiW2018 [234] in 2018), and audio and video (EmotiW2020 [235] in 2020). Within this challenge, hybrid approaches – combining information from both the individual-level emotion and environment context – have shown to result in overall higher accuracy and have become the predominant approach. However, these approaches are mostly tested for images/video, which focused on overt (visible) behavioural features, and not on physiological signals [236].

The literature on collective emotions [237] (see Section 2.4) reports that in group scenarios, individuals have shown spontaneous and unintended similarities in their physiological and behavioural responses [238] – phenomena denoted as physiological synchrony [239], i.e. the inter-dependency or temporal interaction between the physiological data of two or more individuals. Physiological synchrony is closely related to emotional states given that physiological responses are body expressions of emotions (see Section 2.2) [5]. Physiological synchrony has been identified in numerous works through peripheral data across different types of relations, such as parent-child, couples, therapist-client, or social interactions [240, 241]. However, the field of emotion recognition using physiological data has mostly focused on intrapersonal emotion assessment [9, 242], disregarding group-related phenomena such as emotion contagion and physiological synchrony. There is a gap in the use of physiological signals for emotion recognition exploring group interactions, which is the problem addressed in this chapter.

This chapter transitions from individual-level emotion recognition to explore a novel methodology that explores physiological synchrony by performing a weighted average of the groups' emotion class labels to predict the label of an unknown subject. The weights are given by the physiological synchrony between the unknown subject and each member of the group. This approach is explored through the

following research questions:

RQ 1.4: *What synchronization metrics and data representations are most suitable for measuring physiological synchrony for emotion recognition?*

RQ 1.5: *Does the emotion classification accuracy improve with the inclusion of group-level information?*

In summary, this chapter seeks to fill the gap identified in the state of the art regarding group emotion recognition based on peripheral physiological data (in particular, EDA and HRV). Moreover, it explores the potential to improve the accuracy of emotion recognition systems by integrating group context information through a novel metric that incorporates measures of physiological synchrony.

4.2 Background

This section provides an overview of the state of the art on group emotion recognition, the metrics used to measure physiological synchrony, and the datasets that incorporate group physiological data.

4.2.1 Group Emotion Classification

The use of the group context has been successfully employed for emotion recognition in the field of audiovisual content analysis [243], namely through hybrid and top-down approaches [244, 245]. In these methods, in addition to the subjects' facial expressions, information from the global scene is also used, i.e. global features such as skeleton features, and visual attention mechanisms are combined to perform the emotion classification tasks. Similarly, in the field of emotion recognition from speech, group information has also been taken into consideration, namely in dyad conversations [243, 246]. In [243], the authors apply a SVM to predict the listener/speaker emotion by using the facial features and the emotion prediction of the speaker/listener using a SVM over the listener acoustic features. The classification improved when including cross-subject features. In [247], the authors analyze whether the emotional reaction of one individual can be assessed by the emotional response of their partner in a dyad cooperation task, exploring physiological and speech data. The models were trained to predict emotional and non-emotional moments using a linear SVM and a random forest classifier. The results showed that the emotion classification performance increases when combining information from the two subjects. In [246], the authors incorporate time-lagged cosine similarity features on a latent representation from an adapted ResNet architecture performing emotion recognition during dyad conversations using video and audio data. Once again, the experimental results showed that the interpersonal method outperformed the model based on individual features only.

The aforementioned works reinforce the success of integrating group information in emotion assessment, however, these works are based on audio-visual or speech features and focus mostly on dyadic interactions. The study of interpersonal features extracted from physiological data in larger groups was only found in the work by [242]. The authors assess the individual's multi-label categorical emotional state using speech and PPG during group tasks used as input to a transformer encoder block with positional encoding, followed by a bi-LSTM model [242]. The method takes into consideration the group atmosphere given by the aggregation of each group-member score in a self-supervised graph attention network (SuperGAT) relying on cosine similarity, surpassing all the baseline methods in the NTUBA dataset [248] (a dataset from acoustic and PPG within small-group conversation). The literature lacks further validation as it was only tested for a few datasets/use cases (e.g. of three-person small group conversations in [248]). Additionally, the latter work relies on external displays of affection (speech) and does not explore alternative similarity metrics to cosine similarity.

Overall, the literature review shows that the integration of group information in emotion classification tasks can improve classification performance, namely when applied to audiovisual data and in dyadic conversations. Moreover, group emotion recognition based on physiological data is still largely unexplored, and there is a lack of information regarding which synchronization metrics and data representations better describe physiological synchrony, and whether they are replicable across group-related activities (i.e. conversation versus watching a movie).

In this part of the thesis, these challenges are addressed by: 1) Proposing a novel approach integrating group context for emotion recognition using physiological data; 2) Performing a diverse analysis of physiological measures and data representations; and 3) Applying our method across two datasets acquired under different group use cases.

4.2.2 Metrics for Physiological Synchrony

In the literature, a broad range of physiological signals have been used to analyze physiological synchrony including [249]: cardiovascular (e.g. ECG, PPG and HRV-related); respiratory (e.g. respiratory rate, respiratory volume time); electrodermal activity (e.g. EDR, EDL); and thermal (e.g. skin temperature). In this chapter, emphasis is given to signals that can be obtained unobtrusively and continuously from a group, and that have low latency so that they can be applied in real-time and in daily living. Therefore, from the explored signals in the previous chapter, the EDA and cardiovascular activity, namely PPG and the derived HRV were selected, as they have been proven to be insightful views into the subject emotions [9] and allow for continuous and unobtrusive data collection over long periods in daily living. In contrast to sensors like ECG or respiration, which require multiple point contact (e.g. ECG) or are placed in intrusive body locations such as the chest (e.g. respiration).

In a survey analysing over 61 works [249], the authors report high ambiguity in using EDA to assess

physiological synchrony, with many papers identifying synchrony in dyads both using EDL [250], and EDR [251], while others not (in EDL) [252]. Similar findings are obtained for inter-group analysis, with the authors in [253] identifying synchronization between strangers, unlike the authors in [254]. Although the experimental results are conflicting, there is evidence of physiologically related synchrony through EDA measures [241].

Cardiovascular activity can be assessed through HR [255, 256], inter-beat interval [238, 257], or HRV-related features [257, 258]. Similarly to the EDA signal, there is still little consensus in the literature. Physiological synchrony is identified in dyadic conversations through differential equations in [259], audience members and dancers [260] in R-peaks through regression. The opposite is described in [261], where no synchrony was identified in groups of 10 individuals at rest and listening to music. For a more detailed description of the works and findings in the literature about physiological synchrony and related areas, the reader is referred to [249].

Overall, although the literature on physiological synchrony is unclear due to factors such as the diversity of signal sources, analysis metrics, protocol setups, or due to the task itself or activity that was studied, numerous papers confirm the existence of physiological synchrony in EDA and HR [249, 259].

4.2.3 Physiological Datasets for Group Emotion Recognition

The literature [237] reports that group dynamics such as emotion contagion can occur even without face-to-face interactions or non-verbal clues (e.g. social media). Nummenmaa *et al.* [262] describe five types of physiological synchrony in groups: a) Independent units (group sharing physical presence but in independent tasks); b) Externally driven (e.g. group watching a movie); c) Leader-follower/sequential interaction (e.g. meeting); d) Dynamic interaction (e.g. conversation); and e) Group interaction (e.g. group cooperative tasks) [263].

In this chapter, two of these conditions are analysed (dynamic interaction in dyadic conversation, and externally driven by video watching) by using two publicly available datasets. The datasets were selected according to the following requirements: 1) Contain group data (≥ 2 individuals); 2) Contain unobtrusive physiological data, namely EDA and cardiovascular data; and 3) Data is continuously annotated in terms of arousal and valence for long-duration naturalistic scenarios (≥ 10 minutes) to elicit group-emotion related phenomena.

The selected datasets were AMIGOS and K-EmoCon. AMIGOS [87] contains physiological (EDA, ECG, and EEG) and audio-visual (face and full-body video) data collected both in groups (4 individuals) and individual settings. The dataset includes data from 37 subjects watching 4 long videos (> 14 minutes) and 16 short video clips (< 4 minutes). The data was continuously annotated $\in [-1, 1]$ in valence/arousal by 3 experts at 20-second intervals. Only data collected in groups was used in this work. K-EmoCon [137] contains physiological (EDA, PPG, EEG), and audiovisual (face, gesture, speech) data

collected during naturalistic dyadic conversations, namely a debate on social issues. The dataset contains 16 sessions of approximately 10 minutes each. The data was annotated by the debate partner, and an external annotator in a [1, 5] scale at 5-second intervals, both using a valence/arousal space and 18 categorical emotions (see Chapter 7). Herein, external annotations are used to maintain the coherency between the two datasets. A summary description of the datasets is shown in Table 4.1. The labels were divided in binary classification, taking the value of 0 as a threshold to separate the binary classes (of -1 and 1), -1: < 0; 1: >= 0 for valence and -1: <= 0; 1: > 0 for arousal.

Table 4.1: Characterisation of the AMIGOS and K-EmoCon datasets class labels. Nomenclature: Arousal (A); Valence (V); and Quantity (#). Group Homogeneity refers to the % of samples in which all members in the group (except the unknown subject) have equal class labels. A sample size of 20 seconds was considered for both datasets.

Label	AMIGOS		K-EmoCon	
	A	V	A	V
-1	81.12	71.23	80.67	08.08
1	18.88	28.76	19.32	91.92
# samples	43327		3192	
Group Homogeneity	70.50 ± 7.10	69.19 ± 7.40		

4.3 Methods

The proposed approach is denoted as WGS, described in Equation (4.1). In a group context of N subjects, the label of an unknown subject (\hat{y}_s) is given by the weighted average of the remaining group members' labels:

$$\hat{y}_s = \sum_{i=1}^{N-1} W_i y_i \quad (4.1)$$

Where W are the weights denoting the synchronization between the unknown individual s and each of the remaining group-member i . The physiological synchronization is given by:

$$W_i = S(h_i, h_s) \quad (4.2)$$

Where h is the data representation and S is the similarity metric used to obtain the synchronization between two subjects. When the similarity metric returns a distance d instead of correlation, it is converted by $S = \frac{1}{1+d}$. The $\hat{y}_s \in \{-1, 1\}$, consists of a binary problem. When a negative correlation occurs ($S \leq 0$), the subject s is given the opposite label of subject i while, for a positive correlation, the label of the subject i is given. Then, the assigned labels are added and weighted by the synchronization value between the unknown subject and its group members' class labels. When the group consists of a dyad, W is always 1 or -1 , and the unknown subject is given the label of the other member in case of a positive

correlation, or the opposite label for a negative and zero correlation.

4.3.1 Workflow

An overview of the tested methodologies is shown in Figure 4.1. Two alternative inputs were first tested: The *IA* pathway, where the EDA morphology (EDA, EDR and EDL time signals), EDA and HRV hand-crafted features, and images (EDA spectrogram and recurrence plot) are used as input to a neural network where a higher representation is learned to compute the subjects' emotion label in the *MA* or *MB* steps. Or the alternative *IB* path, where no latent representation is used and the data inputs for the neural network (namely, EDA morphology – EDL and EDR; and EDA and HRV hand-crafted features) are used as input for the *MA* or *MB* steps to obtain the emotion prediction.

After the input representation is defined (*IA* or *IB*), two pathways are proposed, *MA* and *MB*, in which the interpersonal and intrapersonal models are tested, respectively. In the *MA* approach (interpersonal model), group synchrony is performed where the subject emotion classification is performed based on the synchronization between the unknown subject sample and the group members' samples (WGS method). The group synchronization method was tested in two pathways: weighted (*CA*) – where a weight is given according to the synchronization value; and average pooling (*CB*) – where a non-weighted average is performed so that the synchronization metrics are not considered.

If the *MB* path is selected (intrapersonal model), two methods were tested: classification by classic machine learning algorithms (*ML*); or the implementation of a deep learning classifier (*DL*), using the feature extraction layer from the *IA* path with the addition of a sigmoid activation function to get a binary arousal/valence classification.

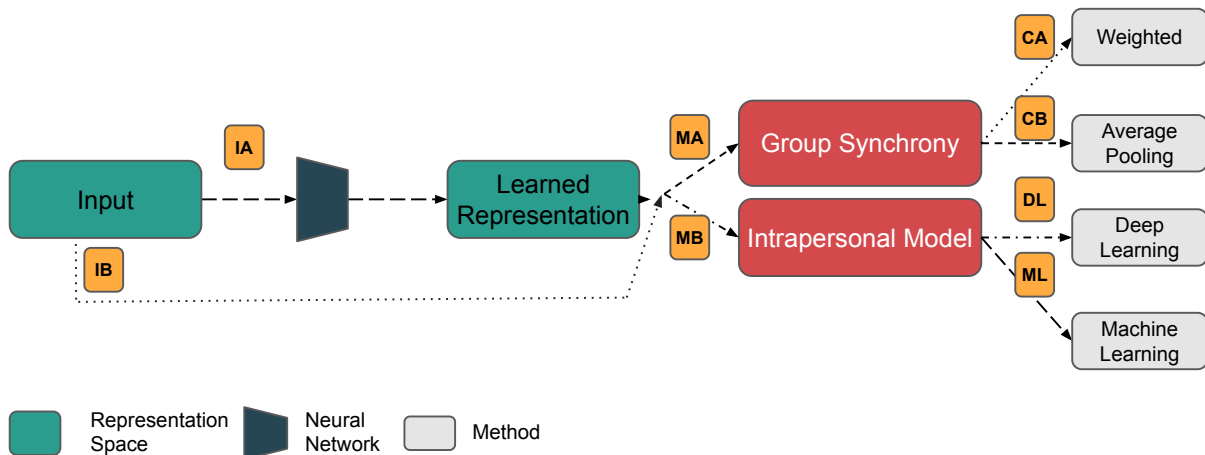


Figure 4.1: Workflow of the tested methodologies. The squares in orange refer to alternative pathways. The *IA* and *IB* paths refer to two alternatives for the input formats: with and without crossing a neural network. The *MA* and *MB* pathways refer to the application of interpersonal and intrapersonal models, respectively. When applying *MA* (interpersonal model), both a weighted (*CA*) and non-weighted (average pooling, *CB*) approach were tested. For the intra-personal model, the application of deep learning (*DL*) and machine learning (*ML*) was tested.

4.3.2 Synchronisation Metrics

To measure physiological synchrony, a set of eight synchronization metrics was considered following two criteria: The first considers the six physiological constructs of physiological synchrony identified in [249] – magnitude, sign, direction, lag, timing, and arousal. Magnitude is determined through Pearson, cosine similarity and Euclidean distance. The sign is determined through Spearman correlation. Direction and lag is determined through Dynamic Time Warping (DTW). Arousal is determined by using EDA data as a correlate for SNS activity. The second criterion considered synchrony metrics found in previous works on the study of physiological synchrony confirming its existence in EDA and HR.

1. **Pearson Correlation** $\in [-1, 1]$: measures the linear correlation between two signals, from negatively correlated to a perfect correlation. Pearson correlation can be interpreted as synchronization magnitude, being one of the most commonly applied metrics in the literature as shown by [264, 265].
2. **Spearman Rank Correlation** $\in [-1, 1]$: analyses the rank-correlation between two signals, from negatively correlated to a perfect correlation, enabling the measurement of the synchrony sign value, i.e. whether signals have the same or opposite dynamics and is applied in [266, 267].
3. **Cosine Similarity** $\in [-1, 1]$: measures the normalized inner product between two signals, and has been used as a magnitude construct of physiological synchrony in [242].
4. **Euclidean Distance** $\in [0, +\infty]$: measures the Pythagorean distance between two signals and has been used as a magnitude construct of synchrony in [268].
5. **Recurrence Plot**, $R^6 \in [0, +\infty]$: The aforementioned metrics are linear, fitting for stationary data with constant mean and variance throughout time. However, physiological data is non-stationary and can show temporal dependency [249]. Recurrence plots allow the characterization of temporal cyclic trends in signals, by filling in the times in which a phase-space trajectory is repeated. Recurrence plots can be found in the literature in [269, 270]. To compare different recurrence plots, six recurrence quantification analysis metrics were extracted: recurrence rate, determinism, average diagonal line length, longest diagonal line length, divergence, and entropy diagonal lines¹.
6. **DTW** $\in [0, +\infty]$: computes the distance between two signals, but instead of calculating the vertical Euclidean distance between the signals, calculates the Euclidean distance across the smallest paths, allowing a temporal synchronization between the signals. The DTW takes into consideration the timing and lag construct of synchronization. The DTW is applied in [271, 272].

¹https://github.com/bmfreis/recurrence_python; Accessed on 20/02/2024

7. **Cross-correlation** $\in [0, +\infty]$: takes into consideration a lag parameter of physiological synchrony to compute the time-shifted correlation between the two signals, using the SciPy correlate function² with mode equal to full, from which the maximum value was obtained to identify the moment of maximum synchrony.
8. **Coherence (Spectral Correlation)** $\in [-1, 1]$: consists of Pearson correlation computed in the frequency domain. The use of spectral metrics is described in the literature to assess synchrony magnitude in [273, 274].

By analysing diverse similarity metrics that explore different characteristics of the data, the work within this chapter further expands the state-of-the-art of emotion recognition using unobtrusive physiological signals.

4.3.3 Data Representation

Given the diversity found in the literature on physiological synchrony, diverse data representation to compute physiological synchrony for emotion recognition were explored:

1. **Signal Morphology**: Corresponds to the cleaned and processed signal in the time domain. For the AMIGOS dataset [87], this work relies on the processed data given by the authors. Regarding the EDA data, the data was further processed by removing existent spikes using the modified Z-score³, the signal was filtered using a Butterworth low-pass filter of 4th order with a cut-off frequency of 5 Hz, and a smoother filter with a window of 0.25 seconds. Afterwards, the signal was normalised ($\frac{y-\mu}{\sigma}$, μ : sample mean; σ : sample standard deviation) for each trial. The ECG signal was filtered using a FIR bandpass filter of 30rd order $\in [3, 45]$ Hz and the R-peaks were computed using the BioSPPy Hamilton segmenter [275]. For the K-EmoCon dataset, the PPG is used to extract the heartbeat peaks. The PPG was filtered using a 4th order Butterworth band-pass filter with 1 – 8Hz cutoff. The heartbeat peaks were extracted using the BioSPPy extractor [276] footnote <https://github.com/scientisst/BioSPPy>; Accessed on 20/02/2024. According to [277], the duration of an emotional response ranges from 0.5 to 4 seconds, reproducing changes in physiological data from 3 to 15 seconds [278]. In both datasets, the data were segmented in 20-second windows with 75% overlap. The EDA data was decomposed into the EDR and EDL components using the cvxEDA library [279].
2. **Image**: Spectrograms are used to collect spectral information and recurrence plots and characterize non-periodic and non-stationary signals. Both have been applied for emotion recognition with state-of-the-art results in the works of [280, 281].

²<https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.correlate.html>; Accessed on 20/02/2024

³<https://towardsdatascience.com/removing-spikes-from-raman-spectra-\8a9fdda0ac22>; Accessed on 20/02/2024

3. **Hand-crafted Features:** A total of 26 features were extracted from the EDA data based on the work by [282], and 76 features from the ECG interbeat intervals. Redundant features (with > 85% correlation) were removed, resulting in 21 for the EDA and 31 features for the HRV. The extracted features are detailed in Table 4.2.

An example with the data representations is shown in Figure 4.2. Figure 4.2 – a, shows the morphological space representation (EDA, EDR and EDL) for one sample (20 seconds) extracted from the AMIGOS dataset. Figure 4.2 – b and c, display a spectrogram and recurrence plot applied on the signal from Figure 4.2 – a, using a Viridis colourmap. Signal morphology (Figure 4.2 – a) and hand-crafted features, were used as the data space input to compute the synchrony between two subjects in the WGS method (*MA* path – interpersonal model). These two spaces, along with the image-based space (Figure 4.2 – b and c) were used as input for a binary arousal/valence classifier, for both a final classification in the *MB* path – intrapersonal model; and to learn a higher level representation in which the WGS was applied (*MA* path – interpersonal model). Table 4.3 displays a summary with the synchronization metrics applied for each data type in the *MA* path – interpersonal model. The data types were divided into morphology and feature-based. The morphology space includes the EDA components (EDR and EDL) through path *IA* and *IB* in Figure 4.1, while the feature-based includes EDA and HRV features used also in both *IA* and *IB* paths in Figure 4.1. The image representation is not included since it was not used to obtain the subjects’ synchronization, only as input for the neural network in the *IA* path in Figure 4.1, with which a feature-based latent representation was learned and then used to calculate the physiological synchronization and the emotion classification label.

Table 4.2: Features extracted from the EDA and RR-interval signals. Features are detailed in Table B 1 and the code for the features is available in the chapter’s repository⁴.

Data	AMIGOS [87]	K-EmoCon [137]
EDA	len pks, pks amp, rise ts, sum pks amp, sum rise ts, sum areas, mean EDA, std EDA, kurtosis EDA, skew EDA, mean 1sder, mean neg 1sder, GSR respEnerg, sum spec, EDR area, spect kurt, mobility, complexity, zero Cross, mfcc kurt, mfcc mean	len pks, pks amp, sum areas, mean EDA, std EDA, kurtosis EDA, skew EDA, mean 1sder, mean neg 1sder, GSR respEnerg, sum spec, EDR area, spect kurt, mobility, zeroCross, mfcc kurt, mfcc mean
RR	nni counter, hr mean, nni diff mean, nni diff min, nn50, pnn20, tinn n, tinn, tri index, fft peak VLF, fft peak LF, fft peak HF, fft abs VLF, fft rel VLF, fft rel LF, fft ratio, lomb peak VLF, lomb peak LF, lomb peak HF, lomb abs VLF, lomb abs LF, lomb rel VLF, lomb rel LF, lomb rel HF, lomb log VLF, ar peak VLF, ar peak LF, ar peak HF, sd ratio, ellipse area, sampen	nni counter, hr mean, nni diff mean, nni diff min, nni diff max, nn50, nn20, tinn n, tinn, tri index, fft peak VLF, fft peak LF, fft peak HF, fft abs VLF, fft abs HF, fft rel VLF, fft rel LF, fft ratio, lomb peak VLF, lomb peak LF, lomb peak HF, lomb abs VLF, lomb abs LF, lomb rel VLF, lomb rel LF, lomb rel HF, lomb log LF, ar peak VLF, ar peak LF, ar peak HF, ar abs VLF, ar rel VLF, sd ratio, sampen

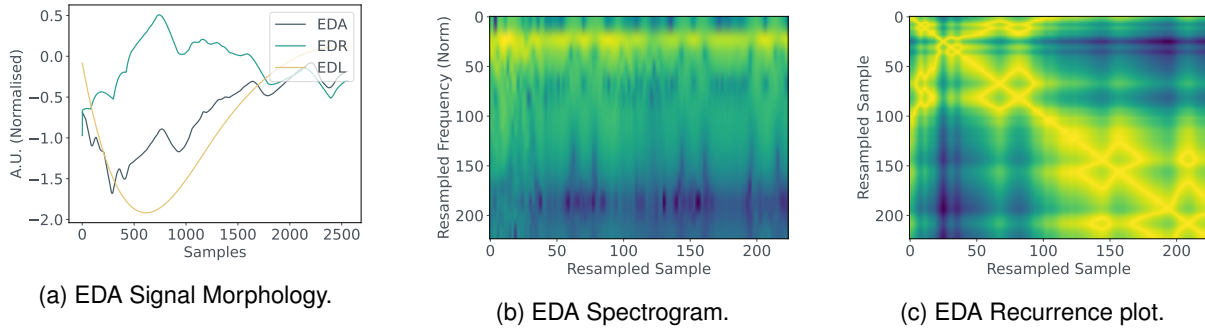


Figure 4.2: Illustration of the data representations: Morphology and Image-based. Data extracted from the AMIGOS dataset. The image data in (b) and (c) were resampled to a 224×224 size to fit the input to the deep learning models.

4.3.4 Classification Models

The state of the art of emotion recognition based on physiological data relies, mostly, on the use of artificial intelligence algorithms incorporating the subject's data with no group context. In this section, the traditional approach found in the state-of-the-art is replicated [9], which does not consider group dynamics in its architecture. This approach is denoted as the intrapersonal methodology, which is used as a benchmark model. Then, the interpersonal model is explored, which incorporates the group context in the emotion classification task. A specific neural network model taking into consideration the characteristics of each data representation is applied:

1. **Signal Morphology:** For one-dimensional data (EDA, EDR, EDL – Morphology representation), the proposed method builds upon the approach from [283], which attained state-of-the-art results for the AMIGOS dataset. The architecture denoted as RTCAN-1D receives as input the three components of the EDA data: EDA, EDR and EDL in three channels. The workflow starts by performing a shallow feature extraction with a convolution layer and batch normalization. Then, a

⁴https://github.com/PatriciaBota/physio_group_emotion_phd/tree/main/group_emotion_recognition; Accessed on 20/02/2024

Table 4.3: Synchrony metrics applied for each data space in the interpersonal methodology. The data representations were divided in morphology space (EDR and EDL signal components used in both *IA* and *IB* paths in Figure 4.1), and feature space (EDA and HRV features in the *IB* path in Figure 4.1; or latent representations from the deep learning models in the *IA* path in Figure 4.1).

Morphology Space	Feature Space
Pearson	Pearson
Spearman	Spearman
Cosine	Cosine
DTW	Euclidean Distance
Euclidean Distance	
Cross-Correlation	
Coherence	
Recurrence Plot	

combination of the three components is performed by an attention layer – Signal Channel Attention (SCA): with two convolution layers, followed by a sigmoid activation, which is multiplied by the attention weight. Temporal similarities are analysed using a non-local attention mechanism relying on an embedded Gaussian kernel as a similarity metric, and a one-dimensional convolution layer followed by average pooling with a kernel of 1 to conduct linear embedding – residual nonlocal temporal attention module. In a third step, an adapted ResNet-18 extracts higher-level features, replacing the two-dimensional convolutions with one-dimensional and simplifying the residual block to perform one-dimensional convolution, batch normalization and a Rectified Linear Unit (ReLU) activation. Lastly, the classification is performed following three fully connected layers, 3 ReLU functions, and a softmax function. A workflow of the architecture can be found in Figure 4.3.

2. **Image:** For the two-dimensional representation (Image – EDA Spectrogram and EDA Recurrence plot), a similar strategy is followed, with a pre-trained ResNet-18 model being applied [284] (Figure 4.4), with state-of-the-art results for the AMIGOS dataset in [280, 281]. The ResNet-18 is based on the addition of residual layers with an identity mapping to the input data, i.e. shortcuts that allow skipping layers. The usage of residual layers has been shown to improve the convergence in deep networks [284].
3. **Hand-crafted Features:** For the hand-crafted features space (EDA and HRV features), the ResNet-18 overall architecture is maintained, with the two-dimensional convolutional layers modified to linear transformation layers.

For all the models, the last layer was changed to set the class number to 1 to perform binary classification. The models were evaluated using LOSO, with one subject left for the test set while the remaining are used in the training set. One subject from the group training set was randomly selected and used as the validation set. The training, testing and validation configuration is the same for both intra- and interpersonal evaluation. The models were developed in Python using the PyTorch library [285], and tuned through the Ray tune library [286] using a grid-search space. Table B 6 shows the values of the hyperparameters used as search space. The Adam optimizer was used as the optimization algorithm.

In addition to deep learning models, classic machine learning algorithms are explored [231]: random forest, SVM, and a naive Bayes classifier, representing a non-linear, linear non-probabilistic, and probabilistic model, respectively. The SVM, and random forest hyperparameters were tuned using a 4-fold CV grid-search.

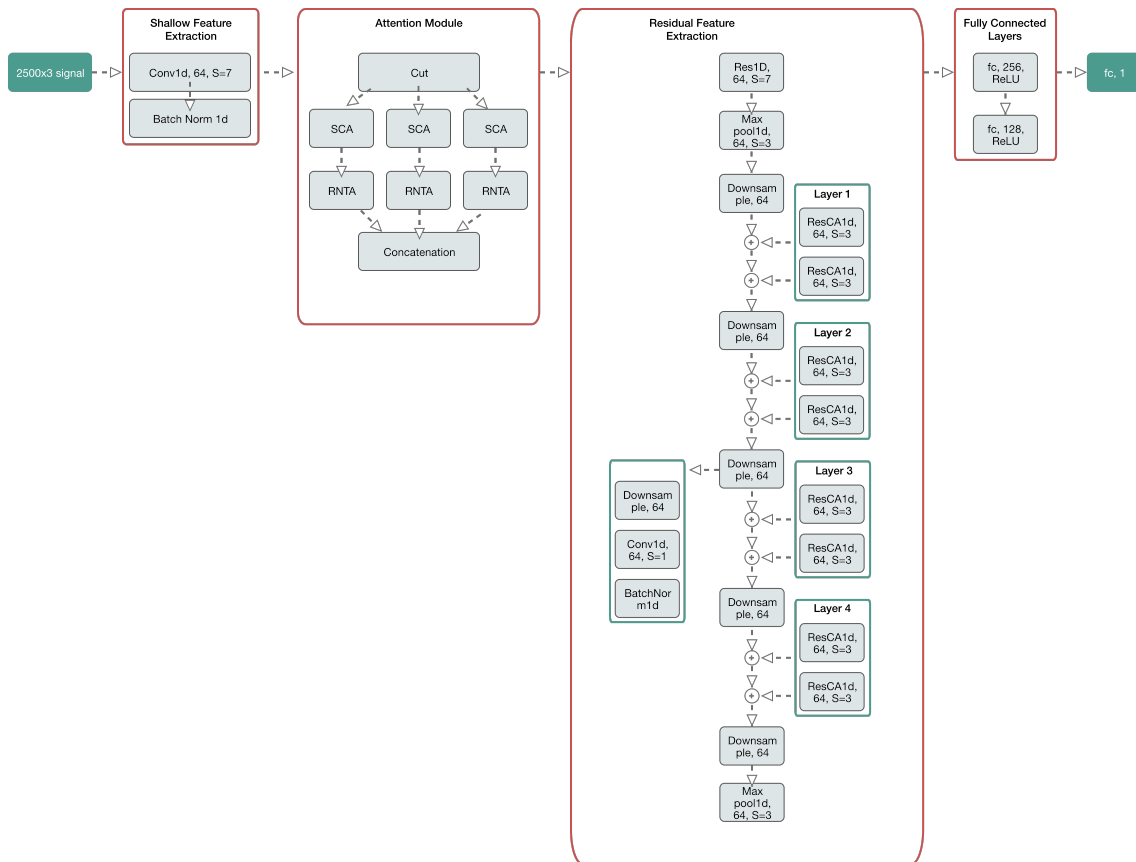


Figure 4.3: The one-dimensional residual temporal and channel attention network (RTCAN-1D) proposed in [283]. Image extracted from [283].

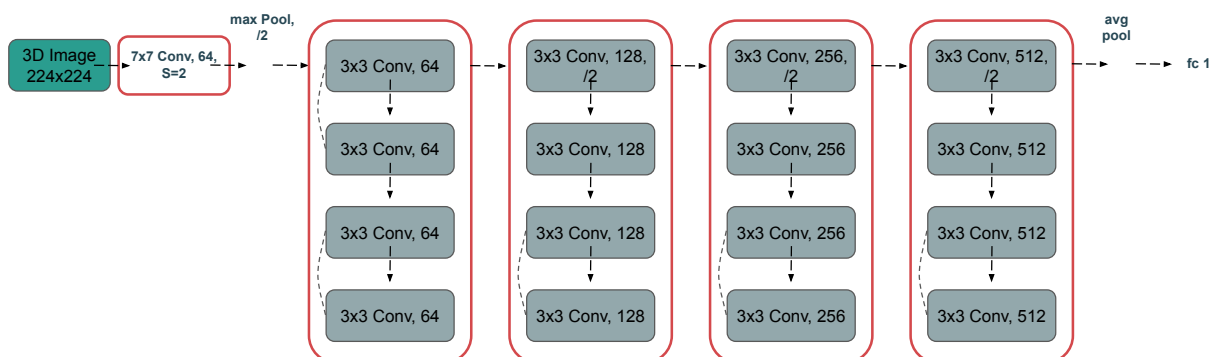


Figure 4.4: ResNet-18 architecture proposed in [284].

4.4 Results

The results are divided into two sub-sections. The first presents the results for the benchmark intrapersonal model (Section 4.4.1) where no group information is embedded into the model architecture. The obtained models are then used to get an additional higher-level space used as input data representation for the interpersonal method (Section 4.4.2).

4.4.1 Intrapersonal Model

For the intrapersonal model, the results are divided by group use case:

AMIGOS Dataset: Table 4.4 shows the intrapersonal model results for the arousal and valence dimensions on the AMIGOS dataset. Due to the heavy data imbalance, two metrics are considered: The Weighted-F1 Score (W-F1), which can be found in the emotion recognition literature, e.g. [277], weights the results by their count value to consider data imbalance. While the Macro-F1 Score (M-F1) performs an unweighted mean of the labels' predictions against their ground truth. Through the rest of the chapter, the analysis is focused on the M-F1 report⁵, while still leaving for observation the W-F1 on the tables, for a more realistic performance in the real world, where data imbalance is expected. Overall, it can be seen that the deep learning models attain similar results or outperform the traditional machine learning models. For the arousal dimension, the best performance is obtained for a fully connected ResNet network using HRV features ($\approx 59.4\%$, M-F1). For the valence dimension, the best performance is obtained when combining EDA and HRV features in a fully connected ResNet ($\approx 66.44\%$, M-F1). The use of images (recurrence plot and spectrogram) or raw data (EDA, EDR, and EDL – Morp.) did not result in improved performance, attaining an F1-score close to random chance in both dimensions. Likewise, for the SVM on EDA data which attained the lowest performance overall ($< 40\%$, M-F1). In addition to the M-F1, the accuracy (which is the most predominant score in the literature) is displayed, although it does not consider data imbalance, and W-F1 (which shows the expected results for an imbalanced distribution). In both metrics, the best performing methodology is above $\approx 69\%$.

K-EmoCon Dataset: Table 4.5 shows that for the K-EmoCon data, the classification performance is overall lower for the arousal dimension, being below random chance. For the arousal dimension, the best performance is obtained for the naive Bayes classifier combining EDA and HRV data. For the valence dimension, the best performance is obtained using the HRV features and a neural network model ($\approx 73.7\%$, M-F1). The deep learning morphology-based and image-based (recurrence plot, spectrogram) methods either outperform traditional machine learning algorithms or attain similarly competitive results. Overall, the best method shows an accuracy greater than 70% and 90% for the arousal and valence dimensions, respectively.

⁵https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html; Accessed on 20/02/2024

Table 4.4: Intrapersonal approach results for the AMIGOS dataset. Nomenclature: Signal morphology space (Morp.); Feature vector (FV); Spectrogram (Spect.); Random Forest (RF); Naive Bayes (NB); Support Vector Machine (SVM); Recurrence Plot (RP); Neural Network (NN).

Data	Model	Acc (%)	W-F1 (%)	M-F1 (%)	Training Time (s)
Arousal					
Morp.	NN	66.73 ± 07.84	66.68 ± 09.67	50.49 ± 05.21	11255.11 ± 191.50
		61.65 ± 06.93	64.16 ± 07.53	50.93 ± 05.84	301.58 ± 273.83
EDA FV	SVM	37.26 ± 10.85	36.70 ± 15.09	34.14 ± 09.42	5.43 ± 3.69
	NB	72.75 ± 06.79	68.92 ± 10.68	50.21 ± 03.82	0.01 ± 0.00
	RF	64.51 ± 10.65	65.26 ± 10.16	51.07 ± 07.73	3.58 ± 3.92
EDA Spect.	NN	70.18 ± 08.73	68.25 ± 10.84	50.66 ± 03.09	1203.70 ± 340.90
EDA RP		70.95 ± 11.32	68.19 ± 11.87	49.37 ± 04.25	2076.95 ± 6608.11
HRV FV	NN	68.51 ± 07.31	70.71 ± 07.81	59.40 ± 08.08	984.82 ± 623.01
	SVM	49.59 ± 11.87	52.45 ± 12.82	45.43 ± 11.99	6.13 ± 1.28
	NB	69.18 ± 17.15	66.76 ± 17.49	51.21 ± 11.47	0.01 ± 0.00
	RF	66.95 ± 09.19	69.20 ± 10.95	58.79 ± 10.78	2.06 ± 1.59
EDA + HRV FV	NN	61.65 ± 6.93	64.16 ± 07.53	50.93 ± 05.84	301.58 ± 273.83
	RF	64.16 ± 11.06	66.40 ± 11.70	56.54 ± 11.39	5.16 ± 4.35
Valence					
Morp.	NN	55.33 ± 08.56	55.06 ± 08.30	49.65 ± 06.47	11197.13 ± 192.68
		55.77 ± 04.51	56.30 ± 04.78	51.78 ± 03.01	283.68 ± 227.95
EDA FV	SVM	41.37 ± 07.15	37.69 ± 09.72	38.59 ± 07.51	5.00 ± 0.98
	NB	61.86 ± 06.59	56.95 ± 08.65	49.02 ± 04.12	0.01 ± 0.00
	RF	52.46 ± 09.84	52.17 ± 10.19	49.13 ± 08.35	1.96 ± 2.69
EDA Spect.	NN	54.55 ± 05.14	55.13 ± 05.67	50.34 ± 02.99	1140.24 ± 409.72
EDA RP		53.69 ± 05.38	54.28 ± 05.54	49.97 ± 03.77	1168.37 ± 436.49
HRV FV	NN	68.05 ± 09.89	67.98 ± 10.30	64.56 ± 10.57	352.97 ± 328.95
	RF	69.04 ± 10.63	68.64 ± 12.10	65.63 ± 11.91	3.30 ± 1.93
	NB	61.85 ± 11.63	59.89 ± 11.42	53.82 ± 08.54	0.02 ± 0.00
	SVM	51.50 ± 10.05	51.40 ± 10.33	49.54 ± 10.07	5.25 ± 1.18
EDA + HRV FV	NN	69.16 ± 09.99	69.36 ± 10.41	66.44 ± 10.05	430.14 ± 319.09
	RF	51.50 ± 10.05	51.40 ± 10.33	49.54 ± 10.07	5.18 ± 1.15

Table 4.5: Intrapersonal approach results for the K-EmoCon dataset.

Data Arousal	Model	Acc (%)	W-F1 (%)	M-F1 (%)	Training Time (s)
Morp.	NN	59.39 ± 22.63	61.59 ± 22.80	45.04 ± 15.44	169.02 ± 134.94
		69.77 ± 14.76	69.10 ± 19.07	45.71 ± 05.24	56.60 ± 83.81
EDA FV	SVM	60.48 ± 19.49	63.49 ± 18.91	45.71 ± 10.00	0.23 ± 0.06
	NB	74.48 ± 17.50	70.97 ± 22.12	46.28 ± 07.22	0.00 ± 0.00
	RF	57.65 ± 18.99	61.27 ± 19.10	44.90 ± 10.84	0.13 ± 0.11
EDA Spect.	NN	69.62 ± 14.77	67.88 ± 20.69	43.87 ± 05.49	608.62 ± 770.74
EDA RP		71.54 ± 15.63	68.93 ± 21.93	47.49 ± 13.53	1962.41 ± 3416.71
HRV FV	NN	68.81 ± 14.14	68.60 ± 19.58	46.30 ± 06.56	59.19 ± 70.02
	SVM	56.20 ± 14.88	61.02 ± 15.58	45.27 ± 09.88	0.40 ± 0.09
	NB	72.76 ± 15.10	71.22 ± 19.47	47.62 ± 07.15	0.00 ± 0.00
	RF	59.64 ± 13.26	63.93 ± 15.50	46.37 ± 08.26	0.32 ± 0.14
EDA + HRV FV	NN	67.37 ± 14.99	68.13 ± 19.91	47.15 ± 06.59	54.93 ± 57.88
	NB	70.68 ± 14.05	70.63 ± 18.32	47.67 ± 05.74	0.00 ± 0.00
Valence					
Morp.	NN	92.86 ± 09.38	90.35 ± 13.23	73.02 ± 27.04	1185.19 ± 875.01
		90.45 ± 10.63	88.95 ± 13.02	62.32 ± 24.80	394.79 ± 370.72
EDA FV	SVM	70.87 ± 18.13	76.83 ± 16.86	43.42 ± 07.47	0.22 ± 0.07
	NB	84.76 ± 10.03	86.83 ± 11.77	47.97 ± 02.84	0.00 ± 0.00
	RF	77.66 ± 15.90	81.79 ± 14.93	50.71 ± 16.88	0.16 ± 0.13
EDA Spect.	NN	93.32 ± 09.34	90.60 ± 13.35	73.14 ± 26.93	479.80 ± 97.12
EDA RP		94.19 ± 07.10	92.96 ± 08.43	71.85 ± 26.40	350.58 ± 118.23
HRV FV	NN	93.36 ± 09.44	90.66 ± 13.41	73.70 ± 26.52	238.49 ± 280.54
	SVM	83.14 ± 09.59	85.89 ± 11.40	48.80 ± 05.40	0.25 ± 0.08
	NB	10.82 ± 10.76	10.66 ± 08.67	10.06 ± 09.79	0.00 ± 0.00
	RF	87.56 ± 10.46	87.94 ± 12.13	55.28 ± 18.90	0.34 ± 0.21
EDA + HRV FV	NN	90.89 ± 09.22	89.36 ± 12.93	55.22 ± 18.98	71.32 ± 81.30
	RF	88.16 ± 12.72	88.02 ± 13.61	60.59 ± 23.18	0.42 ± 0.33

4.4.2 Interpersonal Model

Table 4.6 summarises the main results for the interpersonal approach, namely WGS and average pooling. Table B 2 and Table B 3 (for AMIGOS), and Table B 4 and Table B 5 (for the K-EmoCon dataset), provide the detailed results obtained across all data representations and similarity metrics for the arousal and valence dimensions, respectively.

- **Classification Performance:** analysing Table 4.6, the experimental results show that for the arousal dimension, similarly to what was observed with the intrapersonal method, the WGS applied on the HRV features obtained the best performance ($\approx 72.15\%$, M-F1) using the Euclidean distance to measure physiological synchronization. The use of Euclidean distance on EDL data was also able to maintain the M-F1 average above the 72% mark. For the valence, the best performance was obtained for the EDA features using also the Euclidean distance ($\approx 81.16\%$, M-F1), closely followed by HRV features on the learned representation using cosine similarity ($\approx 81.11\%$, M-F1).

For the K-EmoCon dataset, in the arousal dimension, the cosine similarity on HRV features achieves an equal performance to the use of a non-weighted average pooling ($\approx 52.63\%$, M-F1), with the accuracy ($\approx 83.07\%$) surpassing random chance. For the valence dimension, the best F1-score is obtained for the EDL representation using cross-correlation to measure physiological synchrony ($\approx 65.09\%$, M-F1).

- **Similarity Metric:** Looking at each data representation (Morp. – Morphology (i.e. EDA, EDR, and EDL) vs FV – feature vector vs LR – learned representation) in Table B 2 and Table B 3 for AMIGOS and Table B 4 and Table B 5 for K-EmoCon in Appendix, across dimensions, for the signal morphology spaces (EDR and EDL) the DTW, Euclidean distance, cross-correlation, recurrence plot, and coherence obtain the best performance. Except for feature representations, the use of Pearson, Spearman and cosine similarity often deteriorates the results.

For the AMIGOS dataset, across data representations and dimensions, often cosine similarity shows the lowest Standard Deviation (STD) on the similarity weights (Weight STD), approximating its results to average pooling. The results for the non-weighted group synchronization (average pooling) show similar results with the use of synchronization metrics, outperforming the aforementioned low-quality synchronization metrics (i.e. Pearson, Spearman, cosine similarity). Moreover, for the K-EmoCon dataset, average pooling outperforms the remaining. Additionally, regarding data representations, the EDA and HRV features obtain the most consistent results across synchronization metrics. In terms of accuracy, the best-performing method attains average results $> 80\%$ (arousal and valence) for AMIGOS, and $> 70\%$ (arousal) and 90% (valence) for K-EmoCon.

Table 4.6: Best performing data representation and synchronization metrics for the interpersonal methodology. The results are shown in terms of accuracy (Acc), computation time per sample (Time), and weights standard deviation (Weight STD). The comparable state-of-the-art (SOA) results are shown as well. The best results are shown in bold.

Dataset	Dimension	Data	Similarity Metric	Acc (%)	W-F1 (%)	M-F1 (%)	Time (s)	Weight STD
AMIGOS	Arousal	LR HRV FV	Euclidean Distance	83.07 ± 04.92	82.87 ± 06.05	72.15 ± 10.11	0.007 ± 0.002	0.12 ± 0.05
			Average Pooling	82.65 ± 05.49	82.26 ± 07.05	71.34 ± 10.78	0.010 ± 0.005	0.00 ± 0.00
	Valence	HRV FV – NN EDA FV	Intrapersonal	68.51 ± 07.31	70.71 ± 07.81	59.40 ± 08.08	03.07 ± 01.95	00.06 ± 00.01
			Euclidean Distance	82.80 ± 06.52	83.26 ± 06.09	81.16 ± 07.63	0.004 ± 0.006	00.06 ± 00.01
			Average Pooling	82.70 ± 06.46	83.19 ± 06.01	81.11 ± 07.58	0.008 ± 0.001	00.00 ± 00.00
			Intrapersonal	69.16 ± 09.99	69.36 ± 10.41	66.44 ± 10.05	01.34 ± 00.99	
K-EmoCon	Arousal	HRV FV	Cosine Similarity	70.87 ± 22.50	70.80 ± 22.90	52.63 ± 21.28	0.000 ± 0.000	
			Average Pooling	70.87 ± 22.50	70.80 ± 22.90	52.63 ± 21.28	0.000 ± 0.000	
			Intrapersonal	70.68 ± 14.05	70.63 ± 18.32	47.67 ± 05.74	00.00 ± 00.00	
	Valence	EDA + HRV FV – NB	SOA	55.22		44.86		
			Cross-Correlation (Max)	90.04 ± 08.91	90.16 ± 10.63	65.09 ± 22.55	0.001 ± 0.000	
			Average Pooling	90.04 ± 09.88	90.04 ± 10.94	64.90 ± 22.59	0.000 ± 0.000	
HRV FV – NN	SOA	Intrapersonal	93.36 ± 09.44	90.66 ± 13.41	73.70 ± 26.52	01.88 ± 02.21		
			91.04		87.62			

- **Computational Complexity:** analysing the computation time, overall, a similar order of magnitude (< 0.1 seconds per sample) is obtained across similarity metrics, except for the recurrence plot in the AMIGOS dataset, which shows high prediction times. The recurrence plot computation time is due to the time to compute the recurrence plot and obtain the quantitative analysis features.

4.5 Discussion

This chapter focused on group emotion recognition based on unobtrusive physiological data, evaluating whether physiological synchrony can be identified in diverse group use cases such as dyadic conversations as per the K-EmoCon dataset [137], and group video-watching interaction as per the AMIGOS dataset [87]. Overall, the methodology relying on physiological synchrony holds across datasets attaining competitive results with the state-of-the-art intrapersonal methodology, except for the K-EmoCon dataset on the valence dimension. For the arousal dimension on the K-EmoCon dataset, weighted synchronization on the best-performing metric ($\approx 52.63\%$, M-F1) attains a similar result to performing average pooling ($\approx 52.63\%$, M-F1). Possibly a dyad is not enough to create a group atmosphere with emotion contagion; also group-watching is more prone to emotion contagion and physiological synchronisation than conversation. Another possibility is that synchronization to just one user is more prone to noise than an average over multiple users. Comparing a group of four individuals (AMIGOS) to a dyad (K-EmoCon), the WGS performance decreases in the latter (from $\approx 72.15\%$ arousal, 81.16% valence to $\approx 52.63\%$ arousal, 65.09% valence, M-F1), however, still maintaining score values above random chance.

4.5.1 Synchronisation Metrics and Data Representations

Pertaining RQ 1.4, the focus was on analysing which synchronization metrics and data representations can better measure physiological synchrony for emotion recognition. The experimental results showed that for the AMIGOS dataset, the use of a learned representation on the HRV features attained the highest performance for arousal, and EDA and HRV features for valence. The results for the EDA data on the valence dimension are opposite to what is canonically expected, since EDA is typically associated with arousal [9] and the SNS. The results are possibly due to a high arousal-valence annotation correlation (around 66%), or lower variability in the data (*Weight STD* of 0.06) which could bias for a higher synchronization. For the K-EmoCon dataset, average pooling obtained a similar performance to WGS using cosine similarity on HRV features for arousal, while for the valence space, cross-correlation on the EDL space was the best-performing metric.

Overall, a better performance was observed on the feature space compared to morphology or image-based learned representations (from recurrence plots or spectrograms). Across datasets and dimen-

sions, except for feature representation, the use of Pearson, Spearman and cosine similarity often deteriorates the method's performance. On the whole, the results for the non-weighted group synchronization (average pooling) show similar results with the use of synchronization metrics, outperforming the aforementioned lower performance synchronization metrics (i.e. Pearson, Spearman, cosine). For both datasets, the valence dimension attained a higher classification F1-score when compared to arousal, being in line with what is expected in the literature [9].

4.5.2 Interpersonal Model vs Intrapersonal Model

In what concerns RQ 1.5, the proposed method extends the state of the art by analysing whether emotion recognition can be performed from a subject group members' emotion labels. The experimental results show that the proposed interpersonal methodology outperforms or obtains competitive results comparatively with the state of the art: The authors in [287] report an accuracy of 55.22% (44.86% F1-score) for arousal; and 91.04% (87.62% F1-score) for valence. It should be noticed that the authors rely on additional multi-modal data such as accelerometer, ECG, and skin temperature. For the AMIGOS dataset, to the best of the authors' knowledge, the work herein described is the first to use only the group data (long videos) for a direct comparison of results. The WGS approach surpasses the state of the art in the K-EmoCon on arousal ($\approx 52.63\%$, M-F1 K-EmoCon) and provides novel results for AMIGOS.

The proposed methodology was evaluated against the baseline intrapersonal models, demonstrating that interpersonal models outperform the intrapersonal models for both datasets and dimensions, except for the valence dimension on the K-EmoCon dataset. On the AMIGOS dataset, the intrapersonal model on the arousal and valence dimensions evaluated on M-F1 attained $\approx 59.40\%$ and 66.44% versus $\approx 72.15\%$ and 81.16% for the interpersonal model, respectively. On the K-EmoCon dataset the intrapersonal model on arousal and valence evaluated on M-F1 attained $\approx 47.67\%$, 73.70% versus $\approx 52.63\%$, 65.09% for the interpersonal model.

A limitation that is not controlled for in this work is the existence of pseudosynchrony, i.e., the apparent synchronization between signals even when those signals are not sharing information due to random coincidence. Pseudosynchronisation was identified in the works of [288, 289] where nonverbal synchrony in psychotherapy is analyzed. Due to the phenomena of pseudosynchrony, physiological signals may show higher synchronization than would be expected if the signals were disjoint, leading to a biased (lower accuracy) and less generalization accuracy of WGS than it could be obtained for truly synchronous groups. Thus, pseudosynchrony can explain some of the observed misclassification errors. For the application of hypothesis tests, the creation of pseudogroups (fake synchrony) makes sense since it allows one to compare the pseudogroups to the sample under study, and identify if synchrony exists or not. In this chapter, on the other hand, the goal is to infer if WGS can be applied for emotion recognition classification, or if the phenomena of pseudosynchrony lead to misclassification errors,

making the model unusable in the case of a large classification error.

4.6 Conclusion

The literature on group emotion recognition [149] reports that under certain conditions, the interaction between the group members can lead to emotional dynamics that are distinct from those observed individually, such as emotional contagion and physiological synchrony. This phenomenon can be key for emotion recognition systems, which tend to focus on the data of the subjects individually [242] (intrapersonal methods), missing meaningful context information provided by the group.

This chapter addressed *Obj 1. on the Development of Affective Computing Algorithms*, namely RQ 1.4 and 1.5 on the analysis of synchronization metrics and data representations for group emotion recognition, and the comparison of the proposed interpersonal model (that integrates group information) with the state of the art intrapersonal model.

This chapter expands the state of the art by proposing and evaluating a novel method based on the group members' labels according to their physiological synchrony – the WGS (interpersonal method). To do so, an analysis of synchrony metrics and data representations is performed to analyze which better captures the physiological synchrony interaction in a group setting for emotion recognition systems. Additionally, the method is evaluated under different group sizes (group of 4 versus dyad) and interaction use cases (video-watching versus conversation).

The experimental results show that the WGS integrating group information (interpersonal) can outperform the current state-of-the-art methods based on intrapersonal data (without group context) across datasets and dimensions, except for the valence dimension on the dyads conversation dataset (K-EmoCon). Additionally, the proposed method surpasses the previous works (i.e. [287]) for K-EmoCon on arousal and provides novel comparable results for AMIGOS. Through the analysis of the WGS results on the two datasets, it was possible to conclude that group physiological synchrony contains useful context information for emotion recognition in dyadic conversations and group watching.

This chapter contributes to the field of affective computing by: 1) Introducing physiological synchrony read through physiological signals to emotion recognition; 2) Analysing data representation and physiological synchrony metrics for group emotion recognition; and 3) Improving the accuracy of emotion recognition on group activity tasks over prior work.

Overall, this chapter illustrated how the group context is informative for emotion classification. The next chapter (Chapter 5) will advance the objective of group emotion recognition by evaluating existing platforms for collecting group physiological data, identifying their shortcomings, and proposing a new infrastructure designed to facilitate the gathering of large amounts of emotion-related data in groups and naturalistic settings.

Chapter 5

Data Acquisition

The previous chapter demonstrated that incorporating group emotion into emotion recognition systems can enhance emotion classification. This chapter aims to further the understanding of group emotion analysis by focusing on *Obj 2. on Group-based Physiological Data Collection*, specifically addressing RQ 2.1 and 2.2. It does so through a detailed examination of the current state-of-the-art tools for collecting group data via unobtrusive physiological sensors. The review identified a gap in the field, leading to the creation of an innovative physiological data acquisition system designed specifically for emotion recognition in group and naturalistic settings.

The contents within this chapter were adapted, with permission, from:

- P. Bota, E. Flety, H. P. d. Silva, and A. Fred, “EmotiphAI: a biocybernetic engine for real-time biosignals acquisition in a collective setting,” *Neural Computing and Applications*, vol. 35, no. 8, pp. 5721–5736, 2023

Contents

5.1	Introduction	90
5.2	Background	91
5.3	Methods	92
5.3.1	Wearable	92
5.3.2	Collector	95
5.3.3	User Interface	97
5.3.4	Technical Validation	98
5.4	Results	99
5.4.1	Device Performance Test	99
5.4.2	Infrastructure Test	103
5.4.3	Sampling Period Test	104

5.5	Discussion	105
5.6	Conclusion	106

5.1 Introduction

In recent years, the rapid advancements in microelectronics have facilitated the emergence and proliferation of diverse wearable devices with embedded physiological sensing technologies. These developments have led to the creation of highly integrated sensors that are widely accepted for their ability to be seamlessly integrated into low-cost, non-intrusive, and comfortable wearables, becoming pervasive in users' daily activities.

Simultaneously, the real-world applications of emotion recognition frequently extend to collective environments such as cinemas, theatres, group therapy sessions, museums, and artistic performances. Previous research, as discussed in earlier chapters (Section 2.4), has demonstrated that in group settings—spanning from religious rituals and sports events to anger-related riots—where shared values, focused attention, and behaviours prevail, emotions can undergo synchronization or shared affective states (affective convergence). This leads to a heightened sense of unity within the group, allowing it to act as a cohesive entity. Furthermore, findings from the previous chapter (Chapter 4) have validated that the integration of group emotions into emotion recognition systems boosts emotion classification, underscoring the importance of collective emotional experiences in these applications.

Despite this, a review of current technologies for collecting physiological data reveals a notable gap: there is no available device explicitly tailored for group data acquisition that simultaneously supports a centralised multi-sensor data acquisition at a high sampling rate, offers an open-source framework that can easily be adapted to the problem of emotion recognition, and ensures data privacy. This chapter seeks to introduce and validate a platform for multi-modal individual and collective physiological data collection that addresses these essential criteria.

To do that, the following research questions were addressed:

RQ 2.1: How does the sampling period affect the data loss and transfer quality in multi-device scenarios?

RQ 2.2: To what extent does the network infrastructure influence the maximum number of devices that can collect data without data loss?

Named EmotiphAI, this platform offers a cost-effective and portable solution designed to function on a single-board computer, making it ideal for deployment in various real-world settings. It features a user-friendly interface that allows for the real-time monitoring of the physiological data collection. EmotiphAI distinguishes itself from existing technologies by enabling the simultaneous collection of multimodal data

from multiple sources at high sampling rates in group environments, supporting up to 20 devices at 25 Hz and 10 devices at 60 Hz. EmotiphAI's primary objective is to address the existing shortcomings in the field by offering a powerful tool capable of capturing the complex dynamics of group emotional experiences with precision and ease.

5.2 Background

Table 5.1 illustrates the current wearable devices described in the affective computing literature for data acquisition with embedded PPG and EDA sensors. The detailed analysis of the literature shows that, although there is an extensive list of available devices, only a few were found to effectively possess the characteristics to work in a collective setting. Most devices rely on wireless communication such as Bluetooth or Bluetooth Low Energy (BLE) communication which have a limited number of maximum connections, restricting its applicability to group data acquisition with multiple devices transmitting data simultaneously (Table C 1). Bluetooth also has a physical short-range coverage (1 to 100 meters), depending on the propagation environment, antenna and battery conditions.

An exception from Bluetooth communication is the Xinhua Net FMCI [18] device, which relies on ZigBee communication. Consequently, this thesis starts with the validation of the FMCI device for collective data acquisition, in the scope of a joint project with Xinhua Net, for which the results can be found in [18]. However, critical limitations were identified in the device, namely, it has only one physiological sensor (the EDA), and also a very low sampling rate (1 Hz). It was also identified in [13] that a minimum sampling rate of 10 Hz was required for the EDA signal, and a minimum sampling rate of 50 Hz was required for the PPG. Moreover, the FMCI device and its firmware are not open-source, increasing its cost, and preventing further adaptations to the firmware/hardware that can be required.

A second alternative is the BITalino R-IoT¹, with previously reported uses for data acquisition with 3 to 4 R-IoTs simultaneously at a 200 Hz sampling rate. A higher number of devices may be achieved through the use of several WiFi access points, or decreasing the sampling rate. The R-IoT is a very small and low-cost device with two analogue input ports to which any two sensor modalities can be connected.

A third device identified in the literature is the EmotiBit device [290]. The EmotiBit is a wearable device that measures EDA, PPG, Accelerometer (ACC), Gyroscope (GYR), Magnetometer (MAG), and skin temperature. The EmotiBit allows both Open Sound Control (OSC) by WiFi and Bluetooth communication. However, the EmotiBIT was only commercially available after 3 years (2022). Likewise as for Upmood², only available in 2021, and limited on its measuring signals, only recording PPG.

Within the existing landscape (Table 5.1), the R-IoT was initially selected as the most promising

¹<https://github.com/BITalinoWorld/firmware-bitalino-riot?tab=readme-ov-file>; Accessed on 01/04/2024

²<https://www.upmood.com/group-mood-tracking>; Accessed on 20/02/2024

system for collective data acquisition with multi-modal data at a high sampling rate. Building upon the R-IoT device, in this chapter, the EmotiphAI platform for group data collection is developed and validated, exploring various configurations, such as the number of devices, router points and sampling rates.

5.3 Methods

EmotiphAI is an infrastructure built primarily to facilitate group physiological data collection in the wild, i.e. out of the lab, although it can also be used for individual data acquisition. It consists of a low-cost standalone local infrastructure, a wearable device, and an end-user interface for real-time data visualisation. The EmotiphAI components can be seen in Figure 5.1 and are further described below.

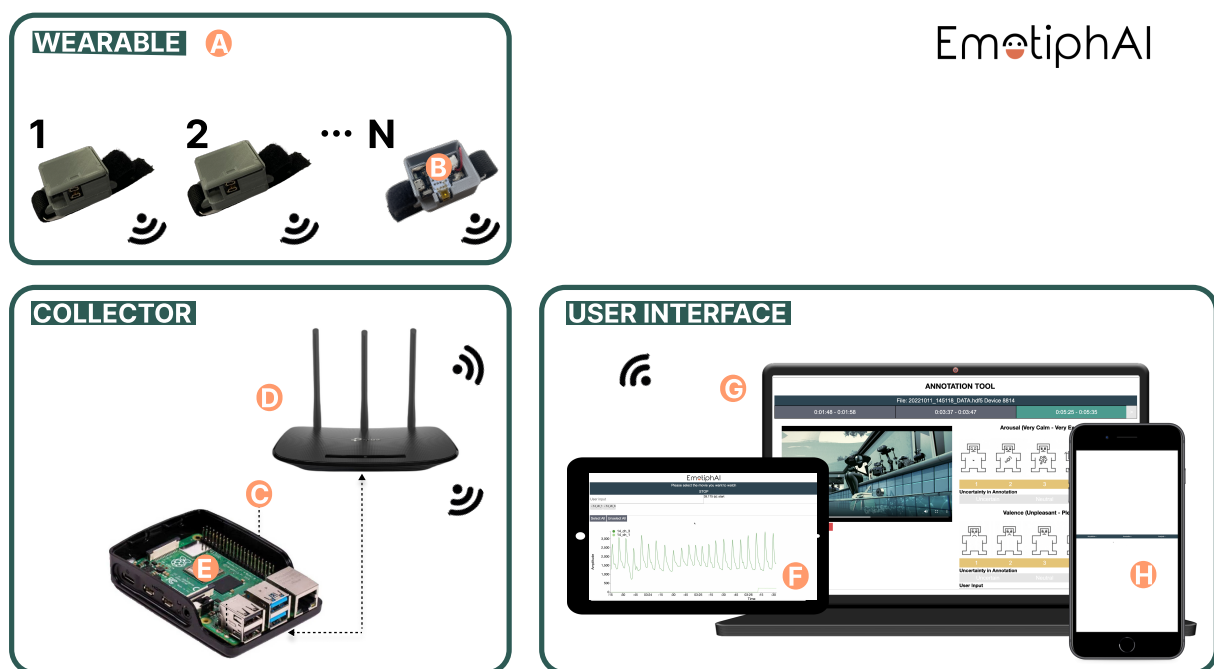


Figure 5.1: EmotiphAI Data Acquisition platform and its components: Collector, Wearable, User Interface. Nomenclature: Wearable (A); Firmware (B); Collector (C); Router (D); EmotiphAI Back-end Software (E); Data Acquisition (F); Data Annotation (G) to be introduced in Chapter 6; and Index Page (H).

5.3.1 Wearable

The literature survey identified the BITalino R-IoT as the most fit device for group physiological data collection. The R-IoT¹ (Figure 5.1 – A), was developed originally by IRCAM³ as low latency, high data rate, and high-resolution device, able to stream gestural information data from multiple performers simultaneously. Its hardware is based on the CC3200 programmable WiFi MCU. In its hardware, it includes a 9-axis digital Inertial Measurement Unit (IMU) sensor (LSM9DS1) (3-axis accelerometer,

³<https://ismm.ircam.fr/riot/>; Accessed on 20/02/2024

Table 5.1: Devices described in the emotion recognition literature for physiological data acquisition of EDA and HR data. The devices are characterized concerning their applicability to research, validation status, form factor, body positioning, battery duration, communication standard, measured signals and, finally, their applicability to a collective acquisition [291]. Nomenclature: Skin temperature (SKT); Respiration (RESP).

Product	Validation	Site	Battery	Comm.	Sensors (Sampling Rate in Hz)	Group
FMCI [18]	Comp. to BITalino [12]	Wrist		Zigbee	EDA (1)	Yes
R-IoT ¹		Non Specific		OSC	EDA, PPG, ACC, GYR, MAG; SR (200)	Yes
EmotiBit [290]		Non-Specific	3.7V	Wifi, Bluetooth	EDA (15), PPG (25), SKT (7.5), ACC (25), Gyr, Mag	Yes
Upmood ²		Wristband	+12 h	Bluetooth	PPG	Yes
BITalino [229]	Comp. to BioPac CE	Non Specific	12-24h	BLE	ECG, EDA, EMG, PPG, RESP (1000)	No
E4 ⁵		Bracelet	+48h	Cloud Storage	EDA (4), PPG (64), SKT (4), ACC (32)	No
MyFeel ⁶	Preliminary study	Wristband	+24h	BLE	EDA, SKT, HR, IMU	No
Oura [292]	Exp. Val.	Ring	2-3 days	Cloud Storage	PPG, ACC, SKT: (250)	No
Bitbrain Ring ⁷		Ring	10h	Bluetooth	PPG, EDA, ACC : (32)	No
CART [293]		Ring		Bluetooth	PPG, ACC	No
Fitbit Charge [294]		Exp. Val.	Smartwatch	5 days	BLE	PPG, ACC
Spire ⁸	Exp. Val.	Clothing Adhesive	1 year	Download or API access	RESP, PPG, ACC	No
Rhythm24 ⁹	FDA clearance	Smartwatch	24h	BLE	EDA, IMP, ACC, Gyr, Mag	No
Heartguide ¹⁰		Smartwatch	2 days	BLE	Blood Pressure	No

⁵<https://www.empatica.com/research/e4/>; Accessed 20/02/2024

⁶<https://www.myfeel.co/hiddenhow-it-workshidden>; Accessed on 20/02/2024

⁷<https://www.bitbrain.com/neurotechnology-products/biosignals/ring>; Accessed on 20/02/2024

⁸<https://www.spirehealth.com>; Accessed on 20/02/2024

⁹<https://www.scosche.com/rhythm24-waterproof-armband-heart-rate-monitor>; Accessed on 20/02/2024

¹⁰<https://omronhealthcare.com/products/heartguide-wearable-blood-pressure-monitor-bp8000m/>; Accessed on 20/02/2024

a 3-axis gyroscope and a 3-axis magnetometer); two 12-bit Analog-to-Digital Converter (ADC) input channels (1.5VMax); one digital input; and one digital output. Its firmware (Figure 5.1 – B) is compatible with the Energia IDE⁴ prototyping platform.

The R-IoT's data is sent using the OSC protocol. The unit being transmitted is an OSC Packet, that is sent from an OSC Client (EmotiphAI device) to an OSC Server (EmotiphAI Collector) as a User Datagram Protocol (UDP) network protocol datagram. For further information regarding the OSC protocol, the reader is referred to¹¹.

Table 5.2: R-IoT Default OSC Message Components.

Component	Description	Range/Units
Accelerometer Axis	3-axis motion detection	[-8; +8] g
Gyroscope Axis	3-axis rotational motion	[-2; +2] %/s
Magnetometer Axis	3-axis magnetic field detection	[-2; +2] gauss
Temperature	Ambient temperature measurement	°K
1-bit Digital Input (GPIO28)	Digital input state	0 / 1
12-bit ADC Inputs (GPIO3 & GPIO4)	Analog to digital conversion	[0; 4095]
Quaternions	Orientation	[-1; 1]
Euler Angles	Node attitude and heading	[-180; 180]°

Additionally, to have a comprehensive understanding of the data being collected for the platform validation, the default R-IoT OSC message was extended to include Received Signal Strength Indicator (RSSI); packet number; and a timestamp. The RSSI enables the measurement of the WiFi link strength; The packet number to detect data loss and out-of-order packets; and the timestamp to determine the precision in the sampling rate as measured by the node timing. Moreover, the packet number is reset to 0 with a 16-bit resolution, resetting to 0 whenever the WiFi connection is lost.

The R-IoT default configuration, namely its sampling rate, or communication port can be changed through a web-based GUI. For further information on how to change the R-IoT configuration, the reader is referred to¹². To encapsulate the R-IoT device, a 3D printed box was designed (Figure 5.2 – A).

Table 5.3: Main Characteristics of the Plux EDA Sensor¹³.

Characteristic	Description
Sampling Rate	100 Hz
Gain	2
Range	0-13 uS (VCC = 3.3 V)
Bandwidth	0-5 Hz
CMRR	100dB
Input Impedance	>1GO hm
Consumption	±0.1 mA

Table 5.4: Main characteristics of the PPG PulseSensor¹⁴.

Characteristic	Description
Supply Voltage	3.3 V or 5 V
PCB Diameter	16 mm
Amplification	330
LED Wavelength	609 nm
Output Signal Range	0 3.3 V (VCC = 3.3 V)
Current Range	< 4 mA

⁴<https://energia.nu>; Accessed on 20/02/2024

¹¹<https://github.com/CNMAT/OpenSoundControl.org?tab=readme-ov-file>; Accessed on 20/02/2024

¹²<https://github.com/BITalinoWorld/riot-python-serverbit>; Accessed on 20/02/2024

¹³<https://support.pluxbiosignals.com/wp-content/uploads/2021/11/eda-sensor-datasheet-revb.pdf>; Accessed on 20/02/2024

¹⁴<https://pulsesensor.com/>; Accessed on 20/02/2024

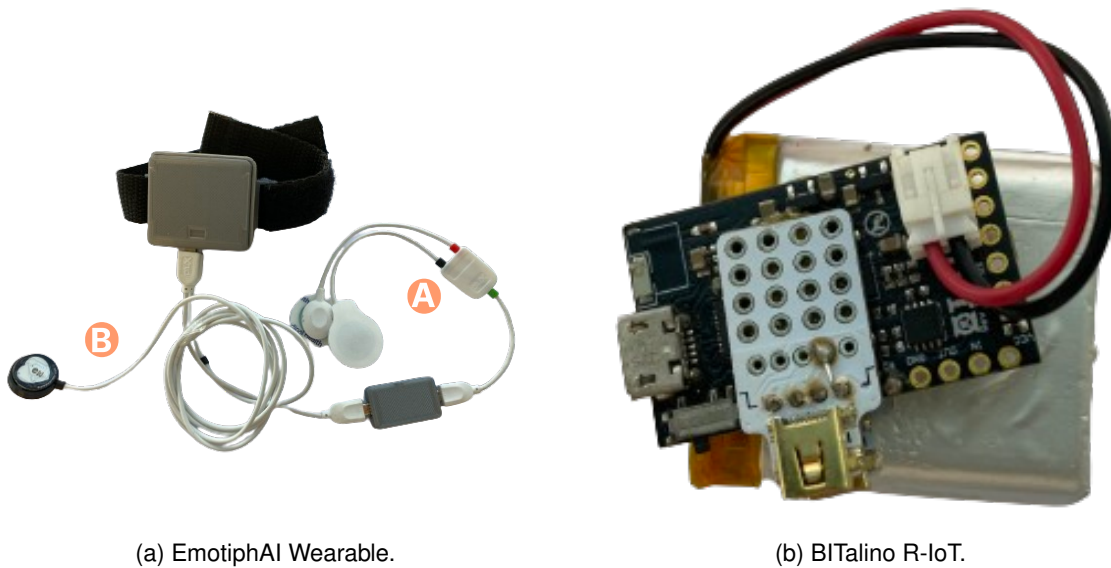


Figure 5.2: EmotiphAI Wearable device with two integrated physiological sensors (the EDA and PPG) built upon the BITalino R-IoT. Nomenclature: EDA sensor (A); and PPG PulseSensor (B).

5.3.2 Collector

A set-top box (Figure 5.1 – C) is used to receive the data from the devices, store it locally and host the graphical user interface. The set-top box is a small single-board computer (for these experiments a Raspberry Pi was used) connected to a router (Figure 5.1 – D) (through an ethernet cable or WiFi) that creates the local network that allows the devices to send the data through.

An overview of the workflow of the EmotiphAI platform is shown in Figure 5.3.

The collector starts the EmotiphAI process at startup, so the user can access its graphic user interface without further configurations. To do so, the user should connect to the WiFi local network created by the router, and access the graphical user interface through a web browser by browsing the single-board computer's local hostname.

The EmotiphAI back-end code (Figure 5.1 – E) is written in Python 3 and the communication to the front-end interface is performed using Flask [295].

For the group data collection software, a thread is created for each device connected to the system. The devices are identified by their ID, which consists of a simplified version of their communication port. Each thread controls a socket that receives the data from the device with a different port per device. Each thread is responsible for receiving the data from the device and storing it locally incrementally as the data arrives. The data is stored in an HDF5 file (Figure 5.4) [296]. Each file contains a group per device with the name of their ID (e.g. 13, 17) (Figure 5.4 – A). The metadata (Figure 5.4 – B) contains information on the data stored: The header attribute contains information on the collected sensors, the sampling rate (in Hz), the video watched during the data collection (optional), and the acquisition starting time in Unix epoch time.

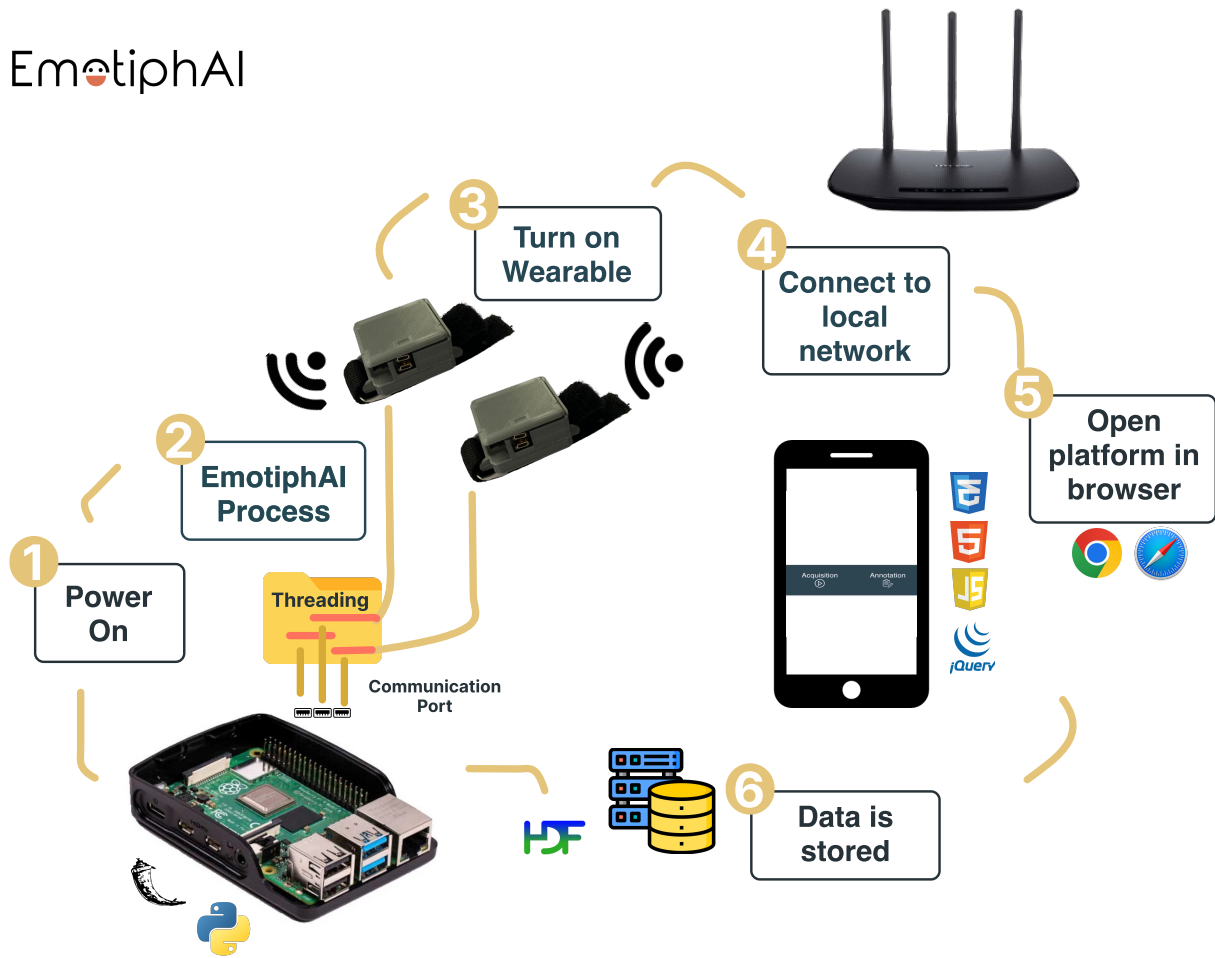


Figure 5.3: EmotiphAI platform workflow.

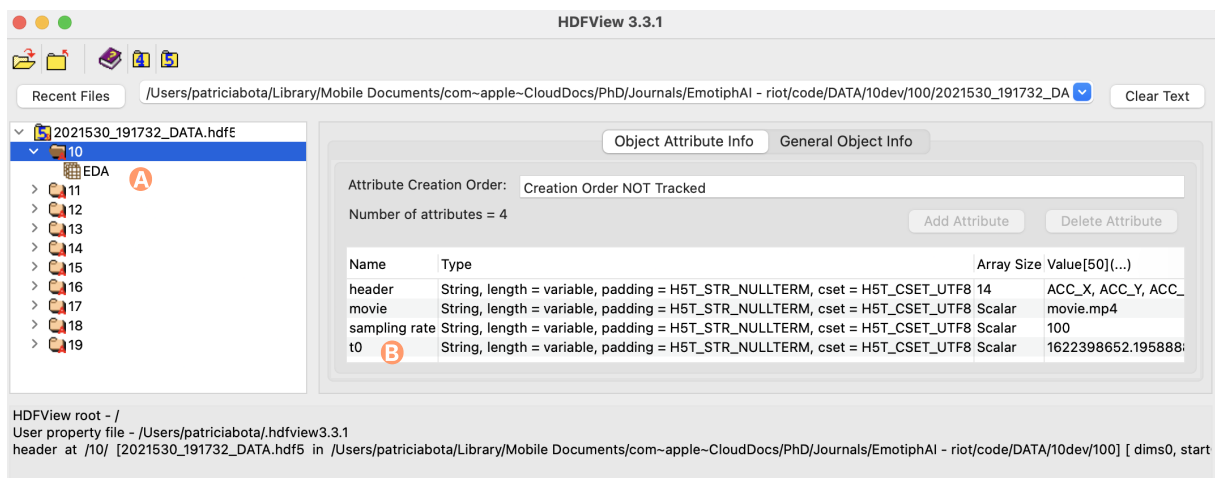


Figure 5.4: Example of EmotiphAI's Data Acquisition HDF5 file. Nomenclature: Data (A); Metadata (B).

5.3.3 User Interface

To facilitate and ensure the correct deployment of the wearable system in the real world, a graphical end-user interface (Figure 5.1 – F, G, H) was developed. The EmotiphAI User Interface contains two main sections, the section for data collection (Figure 5.1 – F) and a section for data annotation (Figure 5.1 – G), which will be the focus in Chapter 6. The Data Acquisition interface allows monitoring the physiological data in real-time (e.g. correct electrode placement), starting and stopping the data acquisition, selecting the devices to be monitored, and annotating the data with metadata (e.g. annotating external events such as phone ringing). The visualization interface is web-based, developed using HTML, CSS and JavaScript.

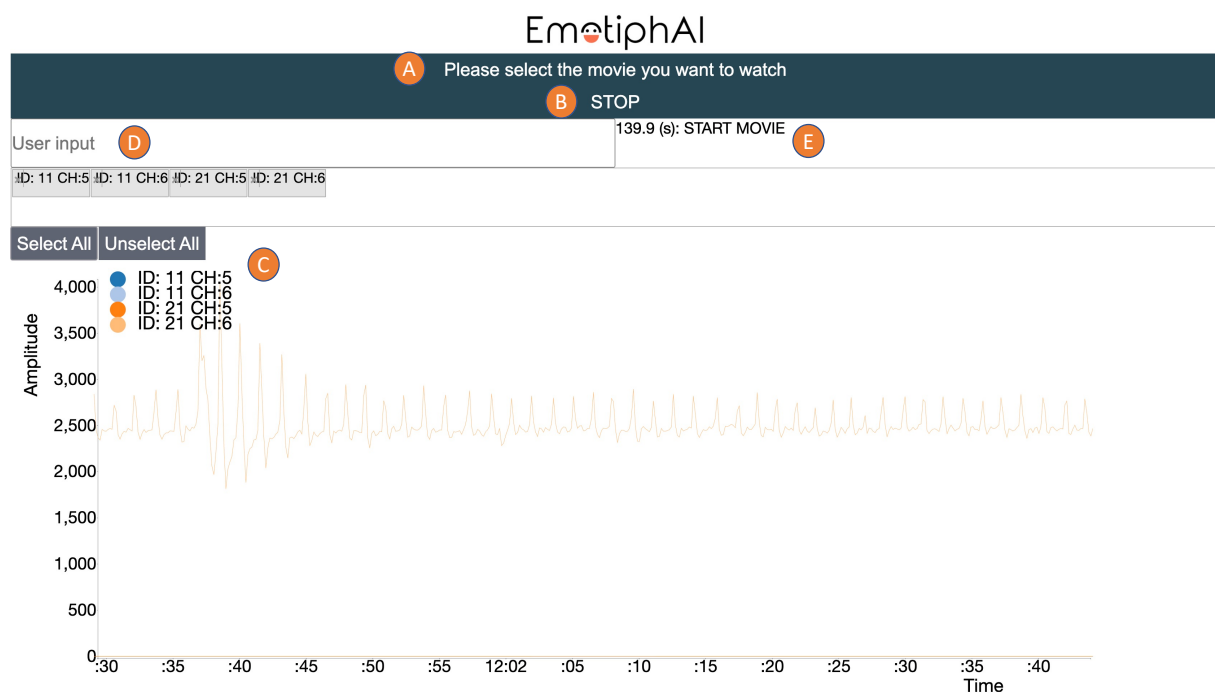


Figure 5.5: EmotiphAI Data Acquisition user interface. Nomenclature: Video selection (A); Start/Stop button (B); Device selection (C); User input box (D); User input history (E).

EmotiphAI has been developed as an integrated platform, with a video player to work as emotion elicitation stimuli. The platform allows the user to select the video they want to play (Figure 5.5 – A). Videos can be stored locally and will be shown in the dropdown menu to provide an emotion elicitation while physiological data is being collected. The selection/visualization of the video is optional. The video and data acquisition begins after the user selects the *START* button (Figure 5.5 – B). The devices connected will be shown in Figure 5.5 – C. The user can observe all the devices connected to the system ("*Select All*"), none ("*Unselect All*"), or one by one by selection in the menu.

The visualization plot has an adaptive zoom to adapt the signals' amplitude.

Moreover, a user input box is available (Figure 5.5 – D), which after submission is shown in the input text history box in Figure 5.5 – E with a timestamp. The data acquisition terminates when the *STOP*

button is pressed (Figure 5.5 – B).

To enable compatibility across different devices (Figure 5.1 – F, G, H), the visualization interface is responsive, automatically adapting the screen size to work both on mobile (i.e. smartphones/tablets) and on the computer. The interface is compatible with predominant web browsers, namely Apple Safari, Google Chrome, Brave and Mozilla Firefox. By hosting the user interface in a web browser, the platform is independent of the operation system (e.g. iOS, Android, MACOS), third-party software, or complex configurations. The platform allows the user to access the interface from any device, as long as it is connected to the EmotiphAI router network.

5.3.4 Technical Validation

A set of comprehensive tests was performed to analyse the performance of the EmotiphAI Collector, namely to identify how many R-IoT devices can transmit data simultaneously and at what sampling rate without data loss and with a homogeneous transmission rate, that will fit the problem of emotion recognition. To do so, a set of tests are performed with different configurations, namely the number of devices, WiFi router and sampling rate.

1. **Device Performance Test:** To test the system performance with 1, 10, and 20 devices for the sampling periods 0.040s (25 Hz), 0.016s (\approx 60 Hz), and 0.005s (200 Hz). A TL-WR940N router was used as an access point. The goal of the experiment was to evaluate if the devices' performance decreases with the increment of the number of devices and determine the maximum sampling rates possible at low data loss for each configuration.
2. **Infrastructure Test:** To compare the use of the TP-Link Wireless N 450Mbps (TL-WR940N) router (Figure 5.6a) to the one recommended by default by the R-IoT manufacturer, which is the TP-Link MR3020 3G/Wi-Fi router (Figure 5.6b). The TL-WR940N has 3 antennas, which allows it to boost the signal, cover more directions, and a larger area in comparison to the MR3020 (which only has one antenna). The technical specifications of the routers are shown in Table C 2. Additionally, the use of two routers was tested, with one working as a bridge router. The goal of the tests was to determine if there was an increase in the network performance in the group acquisitions by increasing the number of antennas or routers.
3. **Sampling Period Test:** To evaluate sampling periods unexplored in the previous tests (40 and 50 Hz). The test was performed using the TL-WR940N router (with 3 antennas).



(a) TP-Link Wireless N 450Mbps (TL-WR940N).



(b) TP-Link MR3020 3G/Wi-Fi.

Figure 5.6: Routers used in the EmotihAI Data Acquisition technical validation tests.

5.4 Results

The EmotiphAI data acquisition platform was tested with different configurations. The evaluated criteria are shown in Table 5.5. Different criteria were evaluated, such as the expected and obtained sampling period to observe if the system was able to maintain the expected sampling period, data loss to observe if the system was able to maintain the data transmission without loss, number of connection resets to observe if the system was losing WiFi connection, time consistency to observe if the data was being transmitted uniformly in time, data shifted to observe if the data was being transmitted in order, duplicated to observe if the data was being duplicated, RSSI to observe the WiFi signal strength, battery to observe the total data collection time, and lastly, the file size to observe the expected data file size. The results are presented in Table 5.6.

5.4.1 Device Performance Test

The performance of the EmotiphAI Data Acquisition platform with 1, 10, and 20 devices was initially analysed. The sampling periods tested were 40 (25 Hz sampling rate), 16 (≈ 62 Hz) and 5 ms (200 Hz). Starting with a sampling period of 5 ms, as it was the maximum sampling period reported for the R-IoT device, then lower sampling rates were tested that considered the minimum sampling period to collect physiological data, namely EDA and PPG. In the work of Silva *et al.*, it was observed that a sampling rate of 50 Hz is enough to capture the PPG signal characteristics at 1 KHz, similarly, a sampling rate of 10 Hz is enough to capture the EDA signal characteristics at 1 KHz [13]. The results for the technical validation tests are shown in Table 5.6 – *1 Dev.*; *10 Dev.*; and *20 Dev.* rows.

1. **Data Loss:** A simplified view of the data loss across the different tested configurations is shown in Figure 5.7. The data loss remains low for 1 device across all the tested sampling periods (40, 16

¹⁵https://github.com/PatriciaBota/physio_group_emotion_phd/tree/main/data_acquisition; Accessed on 20/02/2024

Table 5.5: Metrics and a brief description of the information extracted for the EmotiphAI Data Acquisition technical validation. The code used to obtain the metrics can be found in the chapter's repository¹⁵.

Metric	Description
Number of Devices (Num. Dev.)	Number of devices collecting data simultaneously.
Acquisition Time (Acq. Tm (s))	Duration of the acquisition, with all the devices acquiring data (the time during WiFi reconnection was lost).
Expected sampling period (Exp. SP (s))	Sampling period as defined in the R-IoT configuration page.
Obtained sampling period (Obt. SP (s))	Obtained from the time 1 st derivative. Previous to the derivative computation the time was sorted and only the samples with no packet loss were used to obtain this metric so that the expected sampling period was not biased by data loss.
Loss (Loss (%))	Obtained from the 1 st derivative of the packet number (sorted according to the time), considering as correct samples the number of ones and overflows.
Number of Connection Reset (Rec.)	Number of WiFi re-connections. Counter of the packet number 1 st derivative crosses a defined threshold (-65400, to account for a change from 16-bit resolution packet number to 0).
Time Consistency (Tm C. (%))	Used to determine if the data being transmitted was uniform or streamed in bursts in time (obtained by a normalised unit time histogram of the time vector).
Data Shifted (Shift. (%))	Percentage of data out of order (obtained by the number of negative values in the time vector 1 st derivative).
Duplicated (Duplic. %)	Number of duplicated data samples. Obtained by derivative in time equal to 0.
Incorrect sampling period (Inc. SP (%))	Percentage of the sorted time 1 st derivative where its value does not match the predefined sampling period.
Incorrect sampling period Median (Inc. SP (s))	Median of the sorted time 1 st derivative where its value does not match the predefined sampling period.
Incorrect sampling period Maximum (Max. Inc SP (s))	Maximum value of sorted time 1 st derivative when its value does not match the predefined sampling period.
RSSI (dB)	WiFi Received Signal Strength Indicator.
Battery (hours:minutes:seconds)	Total data collection time.
Size (MB)	Size of the data file in MegaBytes

Table 5.6: Evaluation of the EmotiphAI Data Acquisition platform results, results are presented for different configurations. The following test criteria were evaluated: # Dev.; Exp. SP (s); Acq. Tm (s); Obt. SP (s); Loss (%); Rec.; Tm C. (%); Shif. (%); Duplicated (%); Inc. sampling period (%); Inc. sampling period (s); Max Inc. sampling period (s); RSSI (dB); Battery; Size (MB).

#Dev.	Acq. Tm (s)	Exp. SP (s)	Obt. SP (s)	Loss (%)	Rec.	Tm C. (%)	Shif. (%)	Duplicated (%)	Inc. SP (%)	Inc. SP (s)	Max Inc. SP (s)	RSSI (dB)	Battery	Size (MB)
1 Dev.														
1	34116.738 ± 0.0	0.040 ± 0.0	0.04 ± 0.0	0.447 ± 0.0	0.0 ± 0.0	98.353 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.498 ± 0.0	0.08 ± 0.0	0.16 ± 0.0	-44.5 ± 0.0	9:28:36.78	35.6
1	31718.648 ± 0.0	0.016 ± 0.0	0.016 ± 0.0	0.335 ± 0.0	0.0 ± 0.0	97.712 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.33 ± 0.0	0.032 ± 0.0	0.816 ± 0.0	-50.0 ± 0.0	8:48:38.66	61.4
1	26234.085 ± 0.0	0.005 ± 0.0	0.005 ± 0.0	1.259 ± 0.0	0.0 ± 0.0	98.71 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	36.555 ± 0.0	0.006 ± 0.0	7.795 ± 0.0	-52.0 ± 0.0	7:17:14.09 ± 0:00:00	141.7
10 Dev.														
10	28097.938 ± 0.015	0.040 ± 0.0	0.04 ± 0.0	0.409 ± 0.135	0.0 ± 0.0	90.405 ± 9.369	0.0 ± 0.0	0.0 ± 0.0	0.452 ± 0.151	0.08 ± 0.0	0.12 ± 0.05	-51.75 ± 3.3	9:15:24.05 ± 0:33:56.11	332.4
10	29264.748 ± 0.005	0.016 ± 0.0	0.016 ± 0.0	0.462 ± 0.355	0.0 ± 0.0	98.704 ± 8.794	0.0 ± 0.0	0.0 ± 0.0	0.312 ± 0.201	0.032 ± 0.0	1.152 ± 1.03	-54.0 ± 4.786	8:51:44.61 ± 0:22:27.30	607.4
10	21019.408 ± 0.048	0.005 ± 0.0	0.005 ± 0.0	65.522 ± 17.306	0.0 ± 0.0	26.741 ± 12.962	0.0 ± 0.0	0.0 ± 0.0	62.795 ± 8.366	0.02 ± 0.008	4.935 ± 7.757	-49.75 ± 3.986	6:36:03.29 ± 0:25:34.49	414.5
20 Dev.														
20	27675.945 ± 0.013	0.040 ± 0.0	0.04 ± 0.0	0.358 ± 0.086	0.0 ± 0.0	85.23 ± 26.624	0.0 ± 0.0	0.0 ± 0.0	0.38 ± 0.096	0.08 ± 0.0	0.2 ± 0.07	-51.0 ± 5.459	8:55:32.27 ± 0:31:30.64	655.2
20	21784.863 ± 166.521	0.016 ± 0.0	0.016 ± 0.0	83.827 ± 20.847	0.0 ± 0.0	22.37 ± 20.029	0.0 ± 0.0	0.0 ± 0.0	72.111 ± 21.182	0.096 ± 0.046	96.978 ± 306.25	-52.5 ± 3.675	7:27:08.47 ± 0:57:36.42	305.8
Router														
1	29940.872 ± 0.0	0.005 ± 0.0	0.005 ± 0.0	0.912 ± 0.0	0.0 ± 0.0	85.888 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	44.476 ± 0.0	0.006 ± 0.0	0.066 ± 0.0	-52.5 ± 0.0	8:19:00.88 ± 0:00:00	162.7
10	28785.676 ± 0.006	0.016 ± 0.0	0.016 ± 0.0	0.002 ± 0.003	0.0 ± 0.0	74.964 ± 7.488	0.0 ± 0.0	0.0 ± 0.0	0.002 ± 0.003	0.032 ± 0.0	0.048 ± 0.034	-53.0 ± 3.415	9:03:29.12 ± 0:26:20.76	637.2
20	25872.986 ± 0.046	0.040 ± 0.0	0.04 ± 0.0	39.585 ± 19.223	0.0 ± 0.0	57.636 ± 18.293	0.0 ± 0.0	0.0 ± 0.0	24.71 ± 8.583	0.12 ± 0.046	1.1 ± 185.62	-55.5 ± 5.62	7:57:49.60 ± 0:29:41.88	368.8
20	60543.227 ± 1.46	0.100 ± 0.0	0.1 ± 0.0	0.1 ± 8.695	0.0 ± 1.308	97.568 ± 10.238	0.0 ± 0.0	0.0 ± 0.0	40.705 ± 6.08	0.108 ± 0.001	8.7 ± 3.644	-57.5 ± 4.928	19:01:58.41 ± 1:00:09.92	739.5
Bridge														
10	775.975 ± 0.063	0.005 ± 0.0	0.005 ± 0.0	53.974 ± 12.24	0.0 ± 0.0	42.768 ± 7.225	0.0 ± 0.001	0.0 ± 0.0	30.729 ± 22.332	0.022 ± 0.011	2.191 ± 8.287	-43.5 ± 4.443	0:13:30.95 ± 0:00:15.87	37.9
10	16048.230 ± 0.009	0.010 ± 0.0	0.01 ± 0.0	0.477 ± 0.332	0.0 ± 0.0	98.398 ± 0.657	0.0 ± 0.0	0.0 ± 0.0	0.458 ± 0.307	0.02 ± 0.0	1.186 ± 0.632	-48.5 ± 6.862	7:55:20.74 ± 1:15:28.57	797.3
20	21850.321 ± 0.039	0.016 ± 0.0	0.016 ± 0.0	44.625 ± 27.416	0.0 ± 30.512	54.858 ± 25.665	0.0 ± 0.0	0.0 ± 0.001	31.324 ± 19.419	0.048 ± 0.017	2.823 ± 35.448	-55.25 ± 5.627	6:31:38.39 ± 1:34:23.39	617.5
SP														
1	34190.408 ± 0.0	0.025 ± 0.0	0.03 ± 0.0	0.514 ± 0.0	0.0 ± 0.0	92.315 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	99.985 ± 0.0	0.03 ± 0.0	0.092 ± 0.0	-42.0 ± 0.0	9:29:50.44	39.4
1	27058.240 ± 0.0	0.020 ± 0.0	0.03 ± 0.0	0.473 ± 0.0	0.0 ± 0.0	92.483 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	89.829 ± 0.0	0.03 ± 0.0	0.81 ± 0.0	-44.0 ± 0.0	7:30:58.27	30.7

and 5 ms). As the number of devices is increased to 10, the data loss at a 5 ms sampling period increases to around 65%, while the data loss for the 40 and 16 ms sampling periods remains low. When the number of devices in simultaneous acquisition increases to 20, only the 40 ms sampling period maintains a low data loss (below 0.5%). With 20 devices, a 16 ms sampling period shows on average a loss of 84%.

2. **Sampling Period:** Throughout the different tests it can be seen that the expected sampling period (*Exp. SP (s)* column – Table 5.6) matches the obtained sampling period (*Obt. SP (s)* column – Table 5.6). This column demonstrated a correct data sampling by the device. The incorrect sampling period percentage (*Inc. SP (%)* – Table 5.6) is below 1% for the higher sampling periods (40 and 16 ms). For the higher sampling rates, when the data loss increases, the percentage of samples with incorrect sampling periods increases.
3. **Connection Reset:** The *Rec.* column in Table 5.6 shows that no WiFi reconnections were detected.
4. **Time Consistency:** For 1 device, the obtained values (*Tm C. (%)* column) show that the devices show a homogeneous data transmission at different sampling rates. For 10 devices, as expected by the values of the data loss, the time consistency is above 98% at 16 ms and 40 ms sampling periods, except for the 5 ms acquisition where it decreases to around 30%. A similar behaviour is seen at the 20 devices test, where acceptable time consistency is only obtained for a 40 ms sampling period.
5. **Samples Out of Order & Duplicated:** No samples out of order were observed (*Shif. %* column). Similarly, there are no observed duplicated samples (*Duplicated %* column).
6. **WiFi Signal Quality:** The WiFi connection quality (*RSSI (dB)* column) is inside the recommended range (-30 to -70 dBm). These values were expected since the devices were placed close to the router.
7. **Battery:** Although the battery duration was shown to decrease with the increase of the sampling rate, all the acquisition setups were able to collect data for periods higher than 6h. The R-IoT does not include a low battery indicator and the battery decreases with its collecting data, which could explain outliers in the data (the device stops collecting data and has a lower data collection time than expected).
8. **Storage:** An increase in the HDF5 size was observed as the sampling rate increased. This was expected since a higher sampling rate collects more data than a lower sampling rate.

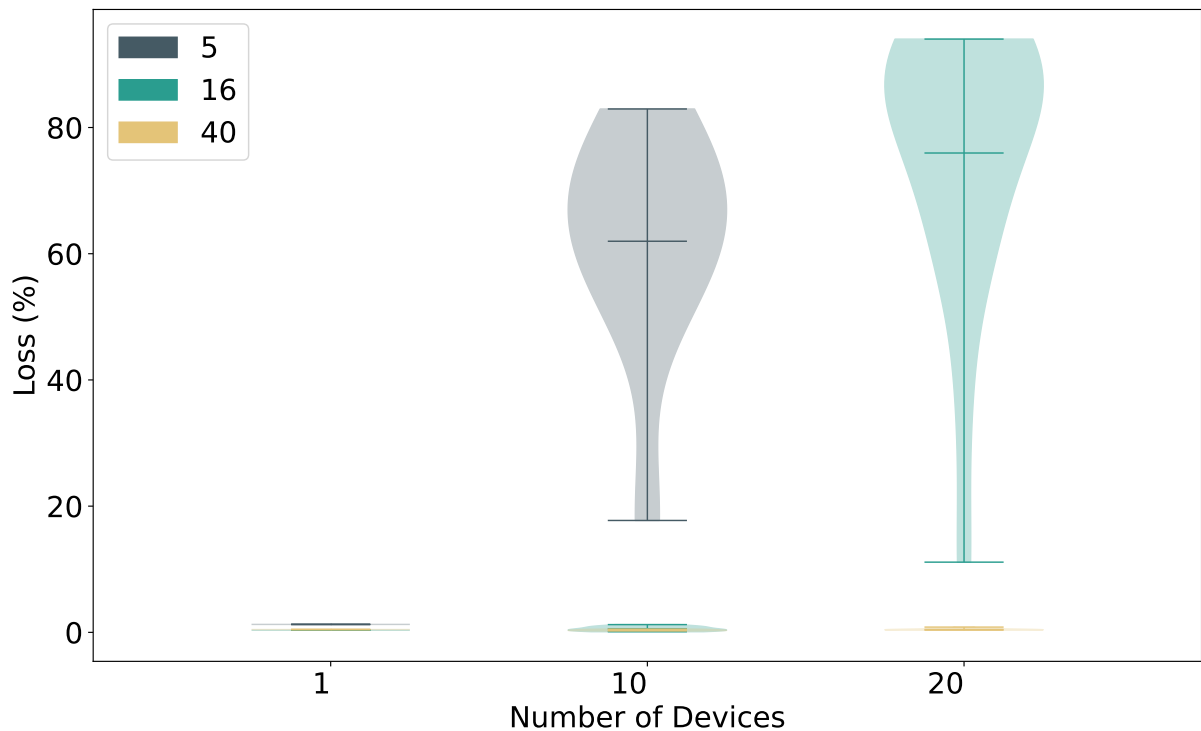


Figure 5.7: Obtained loss for the EmotiphAI Data Acquisition platform Device Performance Test for 1, 10 and 20 devices for the sampling period: 40, 16 and 5 ms.

5.4.2 Infrastructure Test

The last test used a three-antenna router (TP-link MR3020). In this test, the use of a smaller router with only one antenna is tested (TP-link MR3020). The use of 3 antennas has the possibility of boosting the signal and covering a larger area. Additionally, the use of two routers simultaneously is tested, namely: a router TL-WR940N (three antennas) and a bridge repeater TP-link MR3020 (one antenna). The results are shown in Table 5.6 – *Router and Bridge* rows.

1. **Data Loss:** The data loss is illustrated in Figure 5.8. The experimental results show that, for the router with only one antenna, the data loss for one device at a maximum sampling rate (5 ms) is similar (below 1%) to that of the three-antenna router. Likewise, when the number of devices is incremented to 10, the maximum sampling rate is preserved. For 10 devices, it was observed that the use of a repeater router allows to increase the sampling rate to 10 ms (100 Hz) with a low data loss (below 1%). The same is not observed for 20 devices, where the 40 ms (25 Hz) sampling period no longer provides stable data collection for the one internal antenna router, as it did for the three-antenna router. A 20-device acquisition using the TP-link MR3020 is only reliable at 100 ms (10 Hz). Even with the addition of a bridge router, a 20-device data collection was not possible for sampling periods superior to 40 ms (25 Hz).
2. **Time Consistency:** On the whole, even for the cases with low data loss, the data transfer consis-

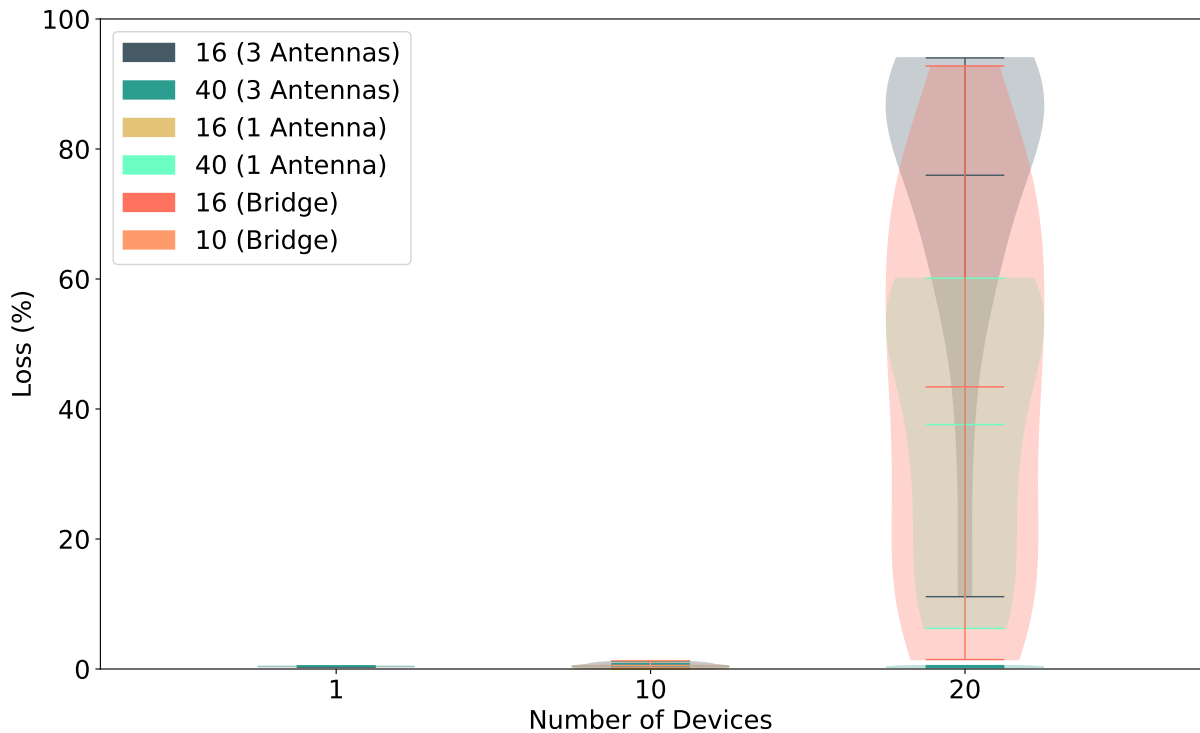


Figure 5.8: Obtained loss for the EmotiphAI Data Acquisition platform Infrastructure test.

tency throughout the data is lower for the TP-link MR3020 than for the TL-WR940N router. The use of the bridge router when the data loss was low was at an acceptable range (above 98%).

3. **Samples Out of Order & Duplicated:** No samples were detected out of order or duplicated.
4. **WiFi Signal Strength:** WiFi connection at the recommended range. The WiFi signal strength did not drop by the use of only one antenna, and neither did it increase significantly by the use of a bridge router.

5.4.3 Sampling Period Test

An atypical behaviour was detected for sampling rates $f_s \in]25, 60[$ Hz which will be explored in this sub-section at a sampling period of 20 ms (40 Hz sampling rate) and 25 ms (50 Hz sampling rate). The device's behaviour at these sampling rates limited the selection of the sampling rate for a collective data collection with 20 devices where lower sampling rates are required to obtain low data loss. The results are shown in Table 5.6 – *SP* rows.

1. **Sampling Period:** Even when collecting data from just one device, the obtained sampling period does not match the expected (i.e. the configured in the R-IoT setup page). When the devices were configured for a sampling period of 25 ms (40 Hz) the obtained sampling period was 30 ms (≈ 33 Hz). Similarly, when a sampling period of 20 ms (50 Hz) was defined, the devices transmitted data at 30 ms.

2. **Data Loss:** The data loss was very low (below 1%). This result indicates that the incorrect sampling period is not caused by data loss. Instead, the cause may be within the device firmware which may be causing the delay during the data transmission.
3. **Time Consistency:** As expected by the low data loss, the data transmission was constant throughout the acquisition (> 92%).
4. **WiFi Signal Strength:** WiFi connection at the recommended range.

For a more comprehensive analysis of the R-IoT device, the authors refer the reader to [12].

5.5 Discussion

The EmotiphAI Data Acquisition platform was evaluated for the collection of data from multiple devices simultaneously across a comprehensive set of tests to address RQ 2.1 on the evaluation of data collection at diverse sampling periods and if they are enough for physiological data collection of EDA and PPG; and RQ 2.2 on the exploration of distinct router configurations.

1. **Device Performance Test:** Addressing RQ 2.1, the R-IoT device was able to reliably collect data with low data loss and a uniform time data transmission in both individual and group settings with 1 device up to 200 Hz (5 ms), 10 devices with a sampling period of 60 Hz (40 ms), and 20 devices up to 25 Hz (40 ms).
2. **Infrastructure Test:** Addressing RQ 2.2, the experimental results show that the use of a three-antenna router does increase the data collection performance for a collective setting. With the use of a one antenna router (TP-link MR3020), 20 devices' data acquisition at 40 ms is no longer possible due to very high data loss (above 39%). The use of a router bridge allowed to increase the sampling rate on a 10-device data collection from 60 Hz (16 ms) to 100 Hz (10 ms).
3. **Sampling Period Test:** The atypical behaviour at]25, 60[Hz sampling rates limited the EmotiphAI data collection for group settings using 20 devices or more since lower sampling rates (< 25 Hz) had to be selected in the previous tests.

Overall, the EmotiphAI platform has shown to be able to perform uniform and reliable data collection across a long period (up to 7h), with very low data loss for 1 device (200 Hz), 10 devices (100 Hz for 2 routers or 60 Hz using 1 router), and 20 devices (25 Hz).

5.6 Conclusion

A look at the literature on physiological sensing for emotion recognition reveals that there are limited technologies available for physiological data collection in group settings. The existing solutions either share their data with 3rd parties, are not open-source, are limited in their physiological sensors, or have a low sampling rate when collecting data from multiple devices simultaneously. This gap has been evidenced by the lack of public datasets for emotion recognition in a group setting.

This chapter addresses *Obj 2. on Group-based Physiological Data Collection*, namely RQ 2.1 and 2.2 on the evaluation of data collection at diverse sampling periods and exploration of distinct network infrastructure configurations.

In this part of the thesis, a platform for multi-modal individual and collective physiological data acquisition is developed. The platform was designated as EmotiphAI and consists of a low-cost standalone local infrastructure run on a single-board computer that reads physiological data through a small, unobtrusive wearable. Additionally, through an end-user interface, it is possible to monitor the data acquisition in real-time. In comparison to the existing solutions. Taking these characteristics, EmotiphAI was developed to be easily transported and deployed in diverse scenarios.

Across a comprehensive set of tests, the experimental results showed that the EmotiphAI platform was able to collect data from multiple devices simultaneously with 1 device up to 200 Hz (5 ms), 10 devices with a sampling period of 60 Hz (40 ms), and 20 devices up to 25 Hz (40 ms). Using a single-antenna router (TP-link MR3020), 20 devices' data acquisition at 40 ms is no longer possible due to very high data loss (above 39%). The use of a router bridge allowed to increase the sampling rate on a 10-device data collection from 60 Hz (16 ms) to 100 Hz (10 ms). Overall, EmotiphAI expands the state of the art by allowing the collection and visualisation of multiple data sources at high sampling rates in a group setting. The platform stores data locally enabling privacy instead of sharing its data with 3rd parties, and its built-in device is small, easily customisable and open-source.

The EmotiphAI platform could be the building block for widespread biosignals data collected from wearables for the creation of large databases for affective computing. Moreover, EmotiphAI has already been used in national and international collaborations, totalling over 400 hours of data collected and over 250 people screened, with installations at the European Commission Joint Research Centre (JRC) Resonances III Art-Science Festival¹⁶, where a system able to access the audience emotional response and dynamically modulate the narrative, audio and visual media of a film correspondingly was deployed (Figure 5.9a); an online theatre performance at the University of Tartu (Figure 5.10a); a live opera at Escola Superior de Música, Artes e Espectáculo (ESMAE) (Figure 5.9b); and the Diferencial cinema sessions at Instituto Superior Técnico (IST) (Figure 5.10b), whose data will be used to create a public dataset further detailed in Chapter 7, among others (see Section 1.3).

¹⁶<https://www.it.pt/News/NewsPost/3541>; Accessed on 20/02/2024

Since emotion recognition algorithms, are heavily dependent on annotated data, the next chapter will present a novel methodology for emotion ground-truth collection tailored for group and naturalistic annotation, so large amounts of data can be collected unobtrusively and quickly.



(a) EmotiphAI deployment at the European Commission JRC Resonances III Art-Science Festival.



(b) EmotiphAI deployment at ESMAE for the live opera performance.

Figure 5.9: EmotiphAI deployments (part 1).



(a) EmotiphAI deployment at the University of Tartu for the online theatre performance.



(b) EmotiphAI deployment at Diferencial Cinema Club.

Figure 5.10: EmotiphAI deployments (part 2).

Chapter 6

Emotion Annotation

In the previous chapter (Chapter 5), the EmotiphAI Data Acquisition platform was introduced. However, the creation of emotion recognition applications requires not only physiological data but also its respective ground truth. The ground truth generally consists of the users' emotional states self-report and is widely used to train emotion recognition algorithms and determine their accuracy.

This chapter focuses on *Obj 3. Emotion Annotation for Naturalistic Settings*, specifically tackling RQ 3.1 to 3.3, by introducing and validating the EmotiphAI Annotator. The EmotiphAI platform facilitates the retrospective annotation of selected moments in extended-duration content, marking a significant advancement in the field of emotion annotation for naturalistic settings.

The contents within this chapter are adapted, with permission, from:

- P. Bota, P. Cesar, A. Fred, and H. Silva, "Exploring retrospective annotation in long-videos for emotion recognition," *IEEE Trans. on Affective Computing*, vol. 15, no. 3, pp. 1–12, 2024

Contents

6.1	Introduction	110
6.2	Background	111
6.2.1	Data Annotation	111
6.2.2	Processes Involved in Emotional Self-Report	113
6.3	Methods	113
6.3.1	Annotation Interface	114
6.3.2	Content Segmentation Method	115
6.3.3	Data Storage	116
6.3.4	Experimental Study Design	117
6.4	Results	118
6.4.1	Usability Test	119

6.4.2	Inter-Subject Agreement	120
6.4.3	Self-Reports Coherence	122
6.4.4	Comparison to Reference Annotations	123
6.4.5	Comparison to Electrodermal Activity	124
6.5	Discussion	126
6.5.1	Usability Test	126
6.5.2	Validity Test	126
6.5.3	Content Segmentation Method	128
6.6	Conclusion	129

6.1 Introduction

The literature offers a broad spectrum of emotion annotation platforms. Their review reveals a reliance on desktop-based systems [141, 297] and real-time annotation [298, 131]. Current methods of emotion annotation, either perform a single post-viewing annotation or perform the concurrent annotation with content viewing, suffering from significant drawbacks. The former may lead to oversimplification and loss of temporal dynamics, while the latter can disrupt the viewing experience, becoming tiring over long sessions, and cumbersome due to the dual tasks of observing and annotating content simultaneously. This often restricts annotations to short video clips in laboratory settings to avoid overwhelming users [9]. However, previous work [6] has shown that emotional responses elicited in laboratory environments can differ from those in naturalistic settings, i.e. closer to real-life experiences. Data collected "in the wild" not only yields higher accuracy but also allows for the accumulation of extensive datasets over prolonged periods.

This research aims to meet this gap by shifting the paradigm of data collection and emotion recognition from laboratory settings to real-world environments, bringing the data collection process closer to natural, everyday experiences. The hypothesis driving this shift suggests that an effective ground-truth emotion data collection platform should facilitate the annotation of long-duration content typical of daily life, ensure an unobtrusive emotional experience to maintain high elicitation intensity and streamline the annotation process to encourage frequent contributions from participants.

This part of the thesis introduces a novel method for the retrospective annotation of long-duration content: the EmotiphAI Annotator, addressing the following research questions:

RQ 3.1: *Are retrospective annotations in long-duration content usable for emotion annotation?*

RQ 3.2: *How do retrospective annotations compare to conventional approaches in long-duration content?*

RQ 3.3: *Which content segmentation method is more suitable for emotion annotation in long-duration content?*

As an alternative to annotating the entire content, the use of a content segmentation algorithm is proposed. This algorithm selects brief clips (10 seconds) based on pre-defined criteria for retrospective emotion annotation. Three content segmentation approaches are explored: 1) EDA-based – SNS-derived segmentation taking into consideration that EDA is a marker of SNS activity (in particular arousal) [119]; 2) Scene-based – time-based using scene boundary detection algorithms described in the state of the art (PySceneDetect¹); and 3) Random selection – used as a null hypothesis. Lastly, the EmotiphAI Annotator is validated for its usability and reliability across a comprehensive set of metrics.

6.2 Background

The accurate annotation of emotional states from collected data underpins the development and validation of emotion recognition models. The emotion annotation process, however, presents unique challenges, including the subjective nature of emotions, the variability in individual emotional responses, and the need for precise, reliable tools to capture these nuances. The annotation platform characteristics can significantly influence the quality of the collected data, the efficiency of the annotation process, and, ultimately, the performance of emotion recognition models trained from the annotated data. Over the years, several emotion annotation platforms have been developed, each with its own set of characteristics designed to address specific aspects of the annotation process.

6.2.1 Data Annotation

Table 6.1 summarizes the predominant annotation platforms in the state of the art, analyzed by their main characteristics.

1. **Device:** The state-of-the-art platforms are divided into mobile and desktop-based. Mobile applications allow out-of-the-lab use and have a smaller form factor [19]. Nonetheless, desktop applications are more predominant in the literature, although they often rely on additional peripherals such as joysticks [131] or wheel mice [302].
2. **Model:** Most platforms rely on the arousal and valence dimensions, either through Russel's Circumplex model [297, 298], or Lang & Bradley's SAM manikins [89, 90, 305]. One exception is Ranktrace [302], in which volunteers rate their tension using a wheel mouse while watching a video. When performing real-time annotation, one dilemma is the use of one or two dimensions. In [300, 302], only one dimension (e.g. to represent either arousal/valence or others), while in most

¹<https://github.com/Breakthrough/PySceneDetect>; Accessed: 20/02/2024

Table 6.1: State of the art emotion annotation platforms. Nomenclature: Real-Time (RT); Continuous (Cont.); Valence (V); Arousal (A).

Platform	Device	Model	Rank	RT	Cont.
FeelTrace [141]		V, A		✓	✓
Gtrace [142]		1d		✓	✓
DARMA [299]		2d		✓	✓
PAGAN [300]	Desktop	1d	✓	✓	✓
AffectRank [297]		V, A	✓	✓	
AffectButton [301]		Image			
RankTrace [302]		1d	✓	✓	✓
NOVA [303]		V,A, Tags	✓		✓
RCEA [298]	Mobile	V, A		✓	✓
EmoWheel [304]		Tags		✓	
EmoteU [19]		V, A	✓	✓	✓
EmotiphAI Annotator	Web	V, A			

works [141, 297, 298] the 2D Circumplex model is used. While one dimension might be insufficient to obtain a comprehensive description of emotion, the use of two dimensions increases the annotation mental workload and is distracting, limiting engagement in the elicitation content.

3. **Rank:** While traditional platforms annotate a magnitude value, rank-based platforms [297, 300, 302] annotate the emotion relative change. In AffectRank [297], the annotation scale is left unbounded, with the annotations returning higher inter-rater agreement when compared to bounded annotation.
4. **Real-time vs Post-hoc | Continuous vs Discrete:** The emotion annotation can be performed in real-time while the subject experiences the emotion, or in post hoc, after the experience retrospectively. Continuous annotation is inherently performed in real-time, allowing the capture of the emotion's temporal dynamics. Discrete methods such as SAM or questionnaires applying basic emotion theory are usually annotated in post hoc [89, 90, 305]. Both real-time and post-hoc annotation show similar mental workload values in [298, 19], while lower values for real-time were reported in [306]. Continuous emotion models are the most common [141, 298, 300].

The review of the state of the art (Table 6.1) shows that the majority of the emotion annotation platforms perform continuous real-time annotation or a single post-hoc discrete annotation of the entire elicitation process. Both are useful to annotate small video clips (< 10 minutes). However, these platforms are not fit to annotate real-life emotional experiences, usually of longer duration, e.g. TV show episodes (\approx 25 to 45 minutes), films (\approx 2 hours), or theatre performances (> 1 hour). In longer-duration elicitation methods, real-time annotation becomes distracting and exhausting, and the use of a single post hoc discrete annotation may not capture detailed information about emotional events, highlighting the need for an annotation method more tailored for naturalistic content.

6.2.2 Processes Involved in Emotional Self-Report

Self-reports can be acquired anywhere, quickly, and regarding any event (e.g., about how the subject felt in the past, how they will feel in the future, or even in hypothetical situations). There is a long tradition of offering retrospective annotations over days, weeks, months [307], after 90 days [308] or even after 1 year [309], without losing their validity and reliability. What changes between different types of annotation are the inherent processes involved in the subject's anchors for the self-report. When the current emotion is not accessible, subjects resort to their memory or their identity-related beliefs and values. Robinson and Clode [307] describe that real-time annotations tend to rely on their direct feelings (experiential knowledge), while for retrospective annotations, episodic memory is used. Although past emotional events cannot be re-experienced, they can be reconstructed by anchoring on relevant thoughts or events. In the work by Ohman [310], the volunteers requested to have access to context over the annotation event; similarly as demonstrated by Siegert *et al.* [311], contextual information improved the inter-rater agreement.

The episodic memory and the ability of the subject to recall past emotions, like any memory, fade with time. The subject will then use situation-specific and identity-related beliefs; the two become relevant when asking for hypothetical, prospective and trait-related emotions. In [312], the authors denoted that daily tiredness and big-five personality traits [313] influence retrospective annotations over a 1-day and 2-week period. Individuals with high neuroticism tend to remember more negative emotions than those reported in momentary ratings, akin to extroverts for positive emotions [308]. Bias is present in all types of emotional self-reports. Age has been shown to influence both momentary and retrospective annotation [314], with older adults being optimistic and rating emotions more positively [309, 315]. Social status can also hinder a true report [5, 316], as well as psychological issues such as alexithymia (the inability to describe one's emotional states) [5, 317].

6.3 Methods

In this part of the thesis, a retrospective annotation tool is introduced for emotion assessment in longer videos – the EmotiphAI Annotator. The platform allows the annotation of longer-duration content by using an algorithm that simplifies the annotation process through the selection of small segments (e.g. 10 seconds) for the user to annotate retrospectively, instead of annotating the entire content (see Figure 6.1). In contrast to the existent platforms, by performing the annotation retrospectively, the EmotiphAI Annotator allows the subject to be fully engaged in the content, not increasing the mental workload during visualization by annotating. The EmotiphAI Annotator is integrated into the EmotiphAI infrastructure (see Chapter 5) as a web-based application that can be used both on mobile and desktop (working across all operating systems) and requires no additional material (e.g. computer mouse or joystick).



Figure 6.1: Content segmentation for retrospective emotion annotation. Moments for annotation are shown in red.

6.3.1 Annotation Interface

Figure 6.2 shows the EmotiphAI Annotator end-user interface for emotion assessment. The users are anonymously identified by their device ID at the top of the page (Figure 6.2 – A) in a dropdown menu. Below, a series of 10-second segments are given for the user to annotate (Figure 6.2 – B). The 10 seconds were selected considering the average duration of an emotional event of arousal 0.5 to 4 seconds and the physiological latency between the stimulus and the reaction (of 1 to 5 seconds) [318, 277]. When the page is loaded, the first segment is automatically selected. In the video player (Figure 6.2 – C), the user can replay and review the video segment. The underlying hypothesis is that, through visualization of the content media, the user can recall their emotional status during the initial visualization of the clip and retrospectively report their emotional state. For the annotation questionnaire, a validated emotion scale is used, the SAM [90] arousal and valence (Figure 6.2 – D), aided by graphics of manikins to express emotional states. Several factors may introduce uncertainty in the annotation, namely, the annotator may not fully understand the concept of valence or arousal, the annotator may not be engaged in the study and respond randomly, and the segment may contain several emotions, among others. For this reason, uncertainty was introduced (Figure 6.2 – E), for the user to detail their level of confidence in the annotation. Lastly, an optional user input box (Figure 6.2 – F) is given for the user to provide additional comments, e.g. indicating an external interference. Once all the information is reported, the next segment will automatically load.

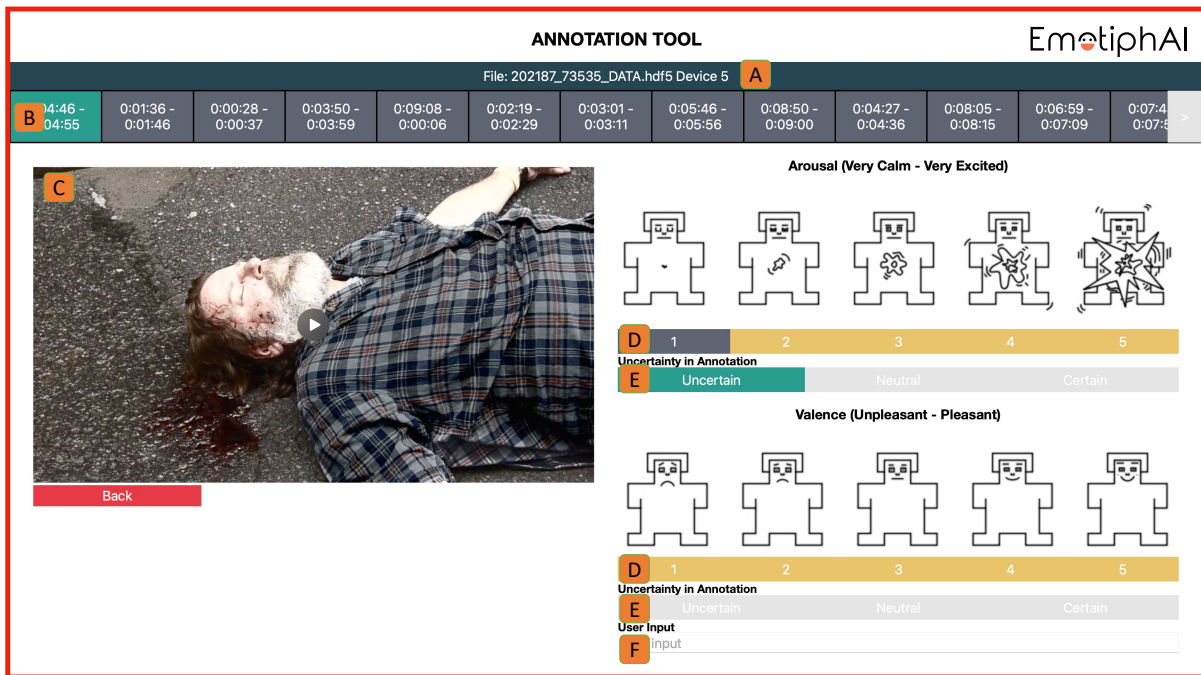


Figure 6.2: EmotiphAI Annotator end-user interface. Nomenclature: File Selection (A); Segment Selection (B); Video Player (C); Arousal/Valence Self-report (D); Uncertainty (E); Text Comment (F).

6.3.2 Content Segmentation Method

The platform contains integrated algorithms (see Figure 6.3) that segment moments of the film for emotion assessment:

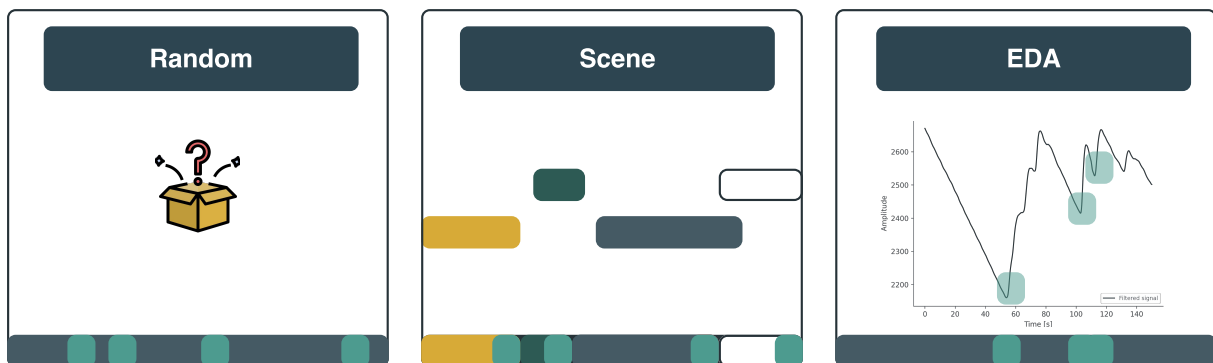


Figure 6.3: EmotiphAI Annotator content segmentation methods for retrospective annotation: Random, Scene-based and EDA-based.

Random

Used for control, it randomly selects 10-second clips from the entire video. Both the number of segments for annotation and their instant are randomly selected using a discrete uniform distribution taking into consideration the length of the video. The number of segments is obtained by a random selection ($[\sqrt{l}/2.5, \sqrt{l}/1.5]$, being l the length of the movie). The random selection method was introduced to compare the obtained results from the aforementioned techniques to random chance. That is, to see if

there is a preferred method to extract meaningful clips for annotation in a long-duration film or, on the other hand, if the entire film is equally meaningful for emotional ground-truth collection.

Scene

Selects for annotation the last 10 seconds of every movie scene, corresponding to a time-sequential annotation. The scenes were identified using existing state-of-the-art methods for scene boundary detection based on image features, namely the PySceneDetect content-aware scene detector, which cuts scenes where the difference between frames exceeds a defined threshold (set to 60 similarly to [319]), empirically found to detect both abrupt transitions and fades to black. The decision to annotate only the last 10 seconds of every scene was taken in order to capture the emotional dynamics of that particular scene, assuming that the emotional intensity or resolution is most pronounced towards the end, offering the most representative snapshot for annotation.

Electrodermal Activity

The EDA-based approach was developed to identify high-intensity SNS-related events. The method selects for annotation, the moments of the video where the subjects' EDA data presented an emotional onset. The data was pre-processed by the application of a low pass 1 Hz cut-off frequency and average window (20 seconds) to remove high-frequency noise and only detect relevant changes in the EDA data. Then, EDA onsets were considered as a minimum on the EDA data followed by an increase of the data higher than a threshold of 0.01% of the EDA maximum [117]. The annotation segment starting time was marked to be 4 seconds before the EDA onset to take into consideration the latency of emotional stimuli (between 1 to 5 seconds [318]). Moreover, the segments were ordered by their event amplitude, following the literature that reports that *"Our most vivid memories tend to be emotional"* [320]. Higher emotional events are easier to recall and, thus are shown first, ensuring that the annotation of the highest-intensity EDA events is made (in a real-world scenario subjects may not complete the rating of all the segments). The EDA fluctuations can occur due to various reasons, including temperature changes, movement, or external distractions. By focusing on higher EDA event values, the aim is to reduce the chances of annotating segments that might be influenced by external factors. Since manual annotation is a resource-intensive process, focusing on segments with high EDA enables the optimization of annotation efforts.

6.3.3 Data Storage

The user's annotations are stored in an HDF5 file. Figure 6.4 illustrates an example file generated by the EmotiphAI Annotator. The data is stored in a hierarchical structure. For each user (e.g. user

with device ID 5) a dataset is created, storing both the physiological data (i.e. Figure 6.4 – A) and the user’s annotations for the segmentation algorithms (Figure 6.4 – B). For each segmentation algorithm and emotion dimension (arousal, valence and uncertainty), the data is gathered in a group. On the bottom (Figure 6.4 – C), the information regarding the metadata for each group is shown, with the number of annotations and a description of the data stored in each column (header attribute). For each annotation, the following information is stored: annotation time (in seconds); annotated segment start time (in seconds); annotated segment end time (in seconds); and self-report value (ranking).

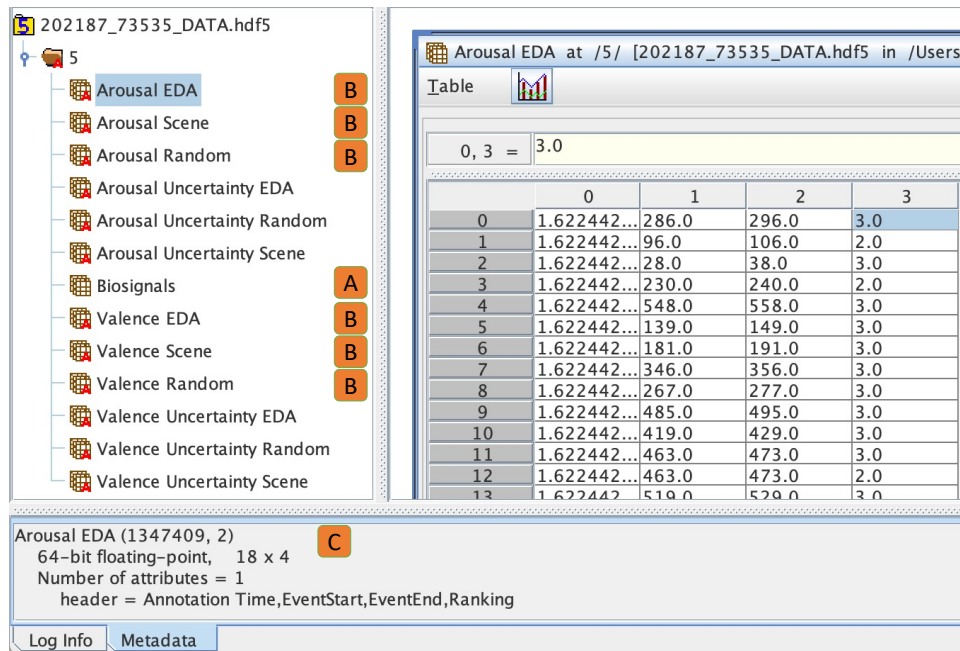


Figure 6.4: EmotiphAI Annotator HDF5 storage file, with the user’s physiological data, annotations and video information. Nomenclature: Physiological data (A); Annotations (B); and Metadata (C).

6.3.4 Experimental Study Design

To analyze the EmotiphAI Annotator usability and annotation reliability, an experimental study was conducted. The study was submitted to and approved by the IST – University of Lisbon Ethics Committee (Ref. n.º 11/2021 Date: 20/04/2021).

- **Procedure:** To ensure the correct deployment of the protocol, an assistant was present during the sessions and guided the subjects across the steps in Figure 6.5. Steps 6 and 7 were repeated three times by every user, one time for each segmentation method.
- **Video Stimuli:** Three videos were extracted from the Continuous LIRIS-ACCEDE dataset [139] (see Chapter 7). The videos were selected based on: 1) Length of the movie – Videos were limited to medium-length videos (i.e. M = 10.72 minutes, STD = 1.09 minutes) to reduce the time and complexity of the protocol; and 2) Valence-Arousal Space – Movies eliciting the four quadrant



Figure 6.5: Experimental protocol deployed to collect the data for the EmotiphAI Annotator technical validation.

areas of the valence-arousal space. The database contains EDA data from 13 subjects watching the videos, and a continuous emotion annotation from 10 volunteers using the GTrace annotation tool (1-dimensional [-1, 1] range) [142].

- **Participants:** A total of 19 participants were recruited to watch three videos. The volunteers' gender and age distribution is shown in Table 6.2. The volunteers varied across videos according to their availability. No participant reported any concerning health condition that would impact the study.
- **Physiological Data:** The physiological data was collected using the EmotiphAI Wearable described in Chapter 5, using pre-gelled self-adhesive Ag/AgCl electrodes connected on the non-dominant hand hypothenar and thenar eminences. Data acquisition was performed at a 60 Hz sampling rate with 12-bit resolution.

Table 6.2 shows the number of subjects, ages and genders per video after subjects with "erroneous" acquisitions were removed. "Erroneous" acquisitions were considered to be those where the data remained at 0 for extended periods, or when there was a problem in the annotation process and the volunteers did not annotate across the three segmentation algorithms.

Table 6.2: Demographic information from EmotiphAI Annotator validation tests: Number of individuals per movie, age and gender distribution (Female (F); and Male (M)).

Movie	Total Indv. (#)	Age (Years)	Gender (% F; M)
After the Rain	18	21.4 ± 1.9	37; 63
Elephant's Dream	12	21.5 ± 1.6	32; 68
Tears of Steel	15	21.9 ± 1.5	21; 79

6.4 Results

The EmotiphAI Annotator's usability and reliability were evaluated across a set of comprehensive tests. Each self-report annotation was extended to a size of 60 seconds. To enhance the comparison of segmentation methods, each video's data was segmented by the physiological data sampling period,

treating the sample points as timestamps. This approach facilitated the temporal alignment and comparison of the outputs from various annotation methods at specific points in time (see Figure 6.6). Only timestamps with annotated segments are used in the comparison analysis, as only these have values to be compared for each method. It is posited that no segment is incorrectly selected, as in the optimal case the entire movie would be annotated. However, as was previously discussed, annotating the entire movie is impractical due to the high burden of broad annotation of long-term content.

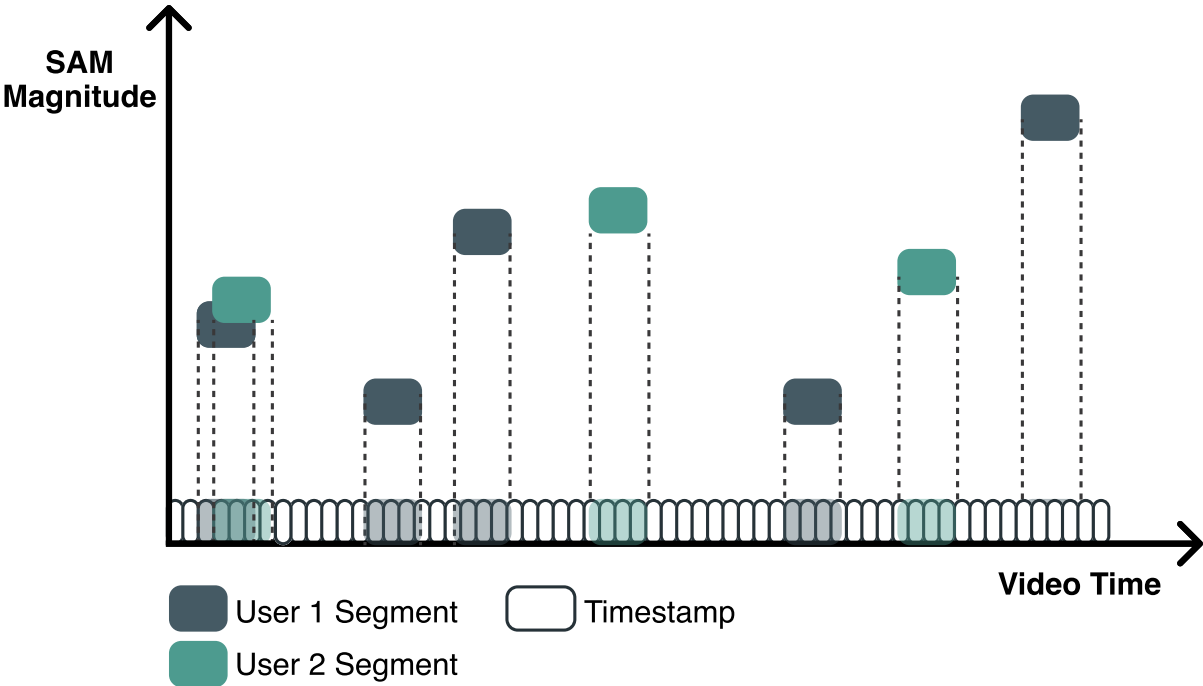


Figure 6.6: Example of the temporal alignment of the EmotiphAI Annotator methods' data by timestamps.

A set of four validity analysis tests were performed to analyze if the EmotiphAI's retrospective annotations are accurate and reliable. The code for the analysis is available at the chapter repository².

6.4.1 Usability Test

Statistics

Table 6.4 presents a summary of the annotations obtained with the content segmentation methods. The results show that, although the number of segments for annotation was lower for the Random algorithm (*Num. Ann. column*), the volunteers took on average a lower time annotating the segments identified using the EDA method (*Ann Time (s) column*). The *Total Ann.* column shows that the Scene algorithm identified the largest number of clips for annotation, while the Random algorithm identified the lowest number. The columns *Ann. Avg Ts.*, *Max Ann. Ts.*, and *Min. Ann. Ts.* provide the average, maximum and minimum number of coinciding annotations performed by the subjects in a given timestamp. A coinciding annotation is observed if there is an annotation in a timestamp for more than one user. The

²https://github.com/PatriciaBota/physio_group_emotion_phd/tree/main/emotion_annotation; Accessed: 20/02/2024

Scene detection algorithm had the largest number of coinciding annotations on average (*Ann. Avg Ts.*), followed by the EDA. Although the EDA detects a lower number of segments for annotation than the Scene method, the maximum number of coinciding annotations in a timestamp is equal. This indicates that the EDA selected at least one common region for annotation across all the subjects, i.e. one section of the movie had a high impact on all the subjects, which was detected in the EDA data. The minimum number of annotations (*Min. Ann. Ts.*) shows that there were regions in which only one annotation (by one subject) was selected using the EDA and Random segmentation. These numbers are expected as the scene detection will select the same segments for every user, the random segmentation should select uncorrelated random segments and the EDA should lead to segments that can be similar (corresponding to high arousal). The different number of annotations across subjects is observed when the subjects forget or skip a clip for annotation.

Usability

Usability and user experience assessment was performed using the System Usability Scale (SUS) [321], and the mental workload was evaluated using the NASA Task-Load Index (NASA-TLX) [322]. A preference question was also introduced. The results are shown in Table 6.3, where the SUS results show a preference for the EDA method with a B⁺ grade score [323].

Mental Workload

For the NASA-TLX, similar values are reported across the three algorithms ($\approx 45\%$), with the lowest reported for the Scene algorithm. The obtained values report that the EmotiphAI Annotator is on the top 40% of the platforms tested by the NASA-TLX questionnaire, corresponding to an average-to-low mental workload. Lastly, the preference question shows that the subjects tended to select as their favourite method the algorithm with a lower number of segments for annotation, i.e. Random method, followed by the EDA and Scene methods.

Table 6.3: EmotiphAI Annotator Usability Questionnaire results.

Method	SUS (%)	NASA-TLX (%)	Preference (%)
Scene	73.50 \pm 1.89	44.74 \pm 0.58	17.67 \pm 7.64
Random	74.75 \pm 3.87	45.21 \pm 1.30	51.00 \pm 3.46
EDA	76.40 \pm 2.11	45.56 \pm 1.48	31.33 \pm 5.51

6.4.2 Inter-Subject Agreement

A common approach in the state of the art to validate self-reports is to calculate the subject's inter-rater agreement, e.g. STD and evaluation error are obtained in [324], correlation, Krippendorff's α ordinal,

Table 6.4: EmotiphAI Annotator evaluation statistics. Nomenclature: Average annotation time between clips (Ann. Time (s)); Number of Annotations per subject (Num. Ann.); Total number of annotations (Total Ann.); Average (Ann. Avg Ts.); Maximum (Max Ann. Ts.); and Minimum number of annotations per timestamp (Min. Ann. Ts.).

Dimension	Method	Ann. Time (s)	Num. Ann. (#)	Total Ann. (#)	Ann. Avg Ts. (#)	Max Ann. Ts. (#)	Min. Ann. Ts. (#)
Arousal	Scene	25.98 ± 0.37	15.41 ± 0.98	468.00 ± 144.87	14.65 ± 2.75	15.00 ± 2.45	14.00 ± 2.45
	Random	27.55 ± 1.02	10.19 ± 0.59	322.33 ± 132.32	09.93 ± 1.43	13.33 ± 1.70	01.00 ± 0.00
	EDA	25.14 ± 0.06	12.84 ± 0.25	400.67 ± 148.87	12.44 ± 2.12	15.00 ± 2.45	01.00 ± 0.00
Valence	Scene	25.63 ± 0.62	15.43 ± 1.00	468.33 ± 144.49	14.65 ± 2.79	15.00 ± 2.45	14.00 ± 2.45
	Random	27.40 ± 0.98	10.17 ± 0.58	321.67 ± 131.94	09.89 ± 1.48	13.33 ± 1.70	01.00 ± 0.00
	EDA	24.90 ± 0.11	12.82 ± 0.24	400.00 ± 148.48	12.43 ± 2.14	15.00 ± 2.45	01.00 ± 0.00

cohen's k , difference to mean, mean absolute difference in [325], Krippendorff's α in [297, 300], and Fleiss' Kappa in [306]. In this work, the metrics applied in [324] are used, where, similarly to this work, naturalistic data was collected and the annotations were performed using the SAM scale. In [324], the authors obtain the inter-subject agreement by analyzing the annotation's STD in a given timestamp, and the evaluation error given by the difference between the average STD in a timestamp and the optimal STD bound [324]. The difference was defined as zero when the STD was below the optimal bound. The optimal and maximum thresholds were obtained using Equation (6.1) & Equation (6.2) [324], respectively, where K is the number of annotators.

$$d_{opt} \leq \frac{1}{2} \sqrt{\frac{\xi}{\xi - 1}}, \text{ with } \xi = \begin{cases} K, & K \text{ even} \\ K + 1, & K \text{ odd} \end{cases} \quad (6.1)$$

$$d_{max} \leq 2 \sqrt{\frac{\xi}{\xi - 1}}, \text{ with } \xi = \begin{cases} K, & K \text{ even} \\ K + 1, & K \text{ odd} \end{cases} \quad (6.2)$$

Table 6.5: EmotiphAI Annotator inter-subjects agreement results, given by the evaluation error (Eval. Error) and STD in a given timestamp. The optimal (d Opt.) and maximum upper bound (d Max.) rate the results performance. Annotations $\in \{1, 5\}$. Nomenclature: Dimension (D); Arousal (A); Valence (V).

D	Method	Eval. Error	STD	d Opt.	d Max.
A	Movie	0.18 ± 0.12	0.69 ± 0.13	0.51 ± 0.0	2.05 ± 0.01
	Random	0.19 ± 0.18	0.71 ± 0.18	0.52 ± 0.0	2.08 ± 0.02
	EDA	0.22 ± 0.17	0.74 ± 0.18	0.52 ± 0.0	2.07 ± 0.01
V	Movie	0.05 ± 0.05	0.55 ± 0.07	0.51 ± 0.0	2.05 ± 0.01
	Random	0.05 ± 0.07	0.54 ± 0.09	0.52 ± 0.0	2.08 ± 0.02
	EDA	0.04 ± 0.06	0.55 ± 0.08	0.52 ± 0.0	2.07 ± 0.01

The experimental results for the EmotiphAI Annotator are shown in Table 6.5. Similar results are obtained for the three segmentation algorithms, with a lower evaluation error obtained for the valence dimension. For both dimensions, the STD is above the optimal upper bound and below the maximum bound. The evaluation error values are lower than half the distances between two discrete SAM numbers, being very low for the valence dimension, and the STD is between one range value.

6.4.3 Self-Reports Coherence

The self-report coherence was analyzed by observing if the three algorithms lead to similar annotations for each movie. The self-reports coherence is obtained by comparing the user's average annotations for each timestamp across the three algorithms using the metrics from [324] detailed in Section 6.4.2 (Table 6.6). The experimental results show an evaluation error of 0, which is obtained due to the STD being below the optimum bound, returning an evaluation error of 0. This leads to the conclusion that for

the three segmentation algorithms, the average annotations are similar across timestamps. The valence dimension shows lower variability; an example of the subjects' average annotations for the 'Elephant's Dream' movie is shown in Figure 6.7. The figure confirms that there is negligible inter-segmentation method variability, with an overall value below a magnitude of 1.

Table 6.6: Similarity between the subjects' average self-reports for the segmentation methods. Annotations $\in \{1, 5\}$.

Dimension	Eval. Error	STD	d Opt.	d Max.
Arousal	0.0 ± 0.0	0.17 ± 0.07	0.58 ± 0.0	2.31 ± 0.0
Valence	0.0 ± 0.0	0.14 ± 0.02	0.58 ± 0.0	2.31 ± 0.0

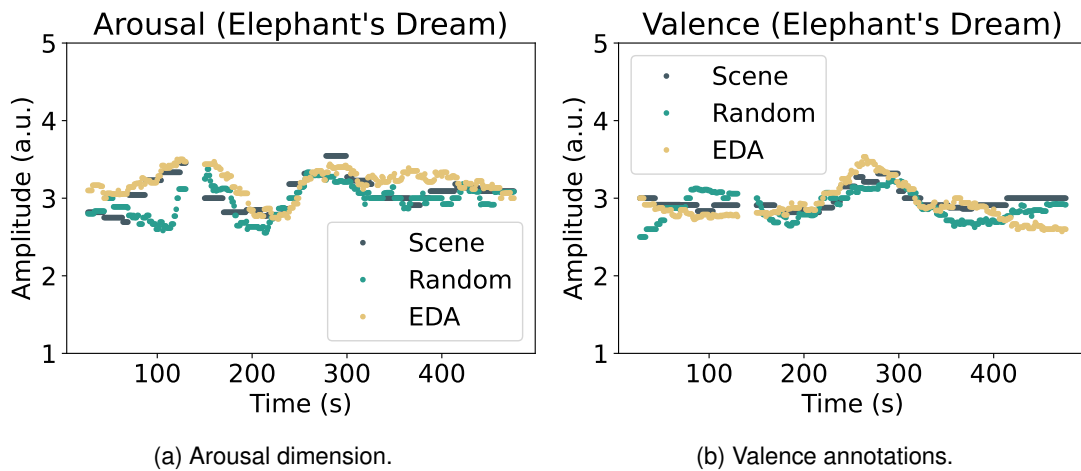


Figure 6.7: EmotiphAI Annotator average annotations for the three segmentation methods.

6.4.4 Comparison to Reference Annotations

The annotations obtained using the EmotiphAI Annotator were compared to the annotations reported in LIRIS-ACCEDE [139], using the metrics by [324] detailed in Section 6.4.2.

Table 6.7: Comparison between the EmotiphAI Annotator and the LIRIS-ACCEDE annotations. Annotations $\in \{1, 5\}$.

D	Method	Eval. Error	STD	d Opt.	d Max.
A	Scene	0.06 ± 0.08	0.65 ± 0.19	0.71 ± 0.0	2.83 ± 0.0
	Random	0.04 ± 0.05	0.64 ± 0.17	0.71 ± 0.0	2.83 ± 0.0
	EDA	0.03 ± 0.04	0.56 ± 0.18	0.71 ± 0.0	2.83 ± 0.0
V	Scene	0.02 ± 0.03	0.49 ± 0.20	0.71 ± 0.0	2.83 ± 0.0
	Random	0.00 ± 0.00	0.62 ± 0.06	0.71 ± 0.0	2.83 ± 0.0
	EDA	0.05 ± 0.07	0.56 ± 0.21	0.71 ± 0.0	2.83 ± 0.0

The LIRIS-ACCEDE baseline was used since it was the only publicly available corpus with physiological data and annotated longer videos in the valence and arousal dimensions identified by the authors. The experimental results in Table 6.7 show that, for each timestamp, the evaluation error is similar across

dimensions and the different segmentation algorithms, with no segmentation method outperforming the remaining. A similar result for the different segmentation algorithms is expected, since the self-reports coherence results in Section 6.4.3 showed that the average annotations across the algorithms are similar. The obtained STD is below the optimal value and the evaluation error is minimal. An example is shown in Figure 6.8, with the comparison of the state-of-the-art ground truth (red) with the obtained annotations using the EmotiphAI algorithms (EDA in green, Random in orange and Scene-based annotations in blue). Although a drift is observed between EmotiphAI Annotator and LIRIS-ACCEDE annotations, on average, the STD value between the two is low (below a magnitude of 0.7 in a [1, 5] scale), and, overall, the annotations time series from both datasets tend to follow the same trend pattern throughout the movie.

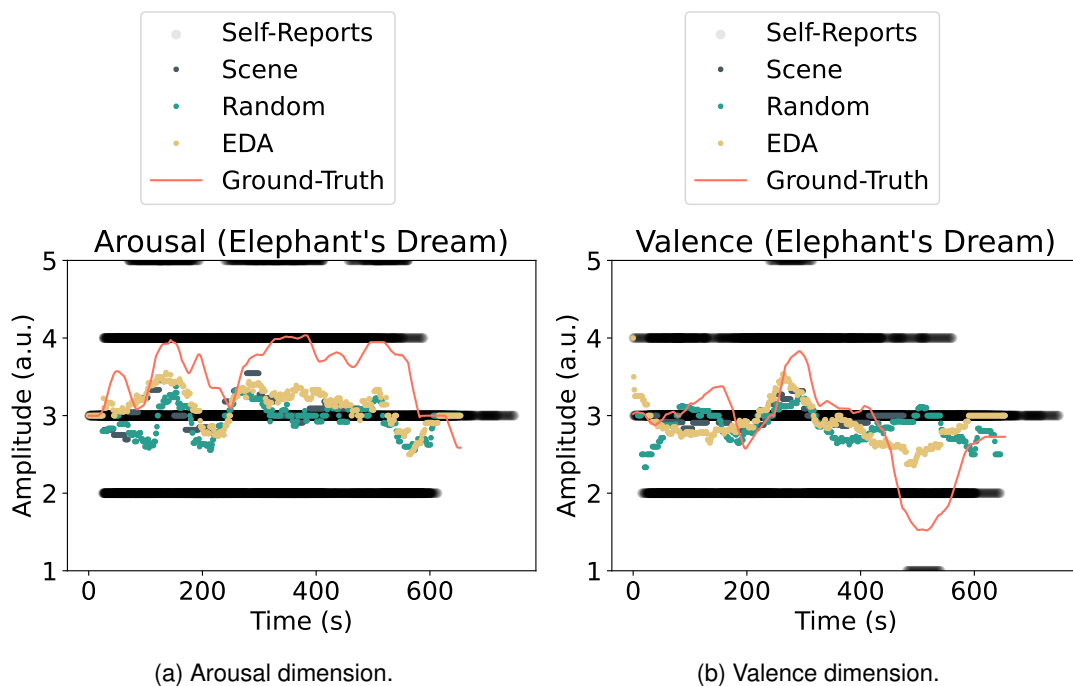


Figure 6.8: Example comparing the LIRIS-ACCEDE ground-truth to the EmotiphAI Annotator.

6.4.5 Comparison to Electrodermal Activity

Considering the EDA as a marker for SNS activity (arousal) [9], an analysis is performed as to whether there is a correlation between the EmotiphAI Annotator arousal self-reports and the collected EDA data. To process the EDA data, the algorithm described in [326] was used, where, similarly to this work, the subjects' EDA given by the MAP was compared to the LIRIS-ACCEDE ground-truth used as a benchmark. The MAP was validated in [327] to contain information regarding the arousal variations of a global audience during a movie. It involves the removal of outlier subjects, a low-pass filter, taking the first derivative, truncation of positive values, and downsampling through a moving average filter. For further information regarding the MAP, the reader is referred to [326, 327].

The experimental results presented in Table 6.8 show that the segmentation correlation results are dependent on the movie. Nevertheless, comparing the three algorithms and throughout the three movies, the EDA method is the only method maintaining a lower-average to above-average correlation between the MAP and the arousal annotations. Poorer results are obtained for the video "Elephant's Dream", which may be explained by the low self-reports variability ($M = 3.33$, $STD = 0.10$). The data from the "After the Rain" and "Tears of Steel" movies is shown in Figure 6.9, where it can be seen that an increase in the annotation value is followed by an increase in the EDA MAP, and vice-versa.

In Table 6.9 the EmotiphAI Annotator's best results are evaluated comparatively to the results obtained by the state-of-the-art [326]. Overall, EmotiphAI Annotator obtains a higher correlation between the arousal self-reports and the MAP, outperforming the state-of-the-art annotations for two of the movies ("Tears of Steel" and "After the Rain").

Table 6.8: EmotiphAI Annotator methods' arousal annotations correlation to the MAP given by the Pearson (Pearson C) and Spearman correlation (Spearman C) in a [0, 1] scale.

Movie	Dimension	Pearson C	Spearman C
Tears of Steel	Scene	0.77	0.73
	Random	0.58	0.65
	EDA	0.58	0.55
After the Rain	Scene	0.62	0.31
	Random	0.32	0.18
	EDA	0.80	0.65
Elephant's Dream	Scene	-0.10	-0.18
	Random	-0.21	-0.20
	EDA	0.26	0.25

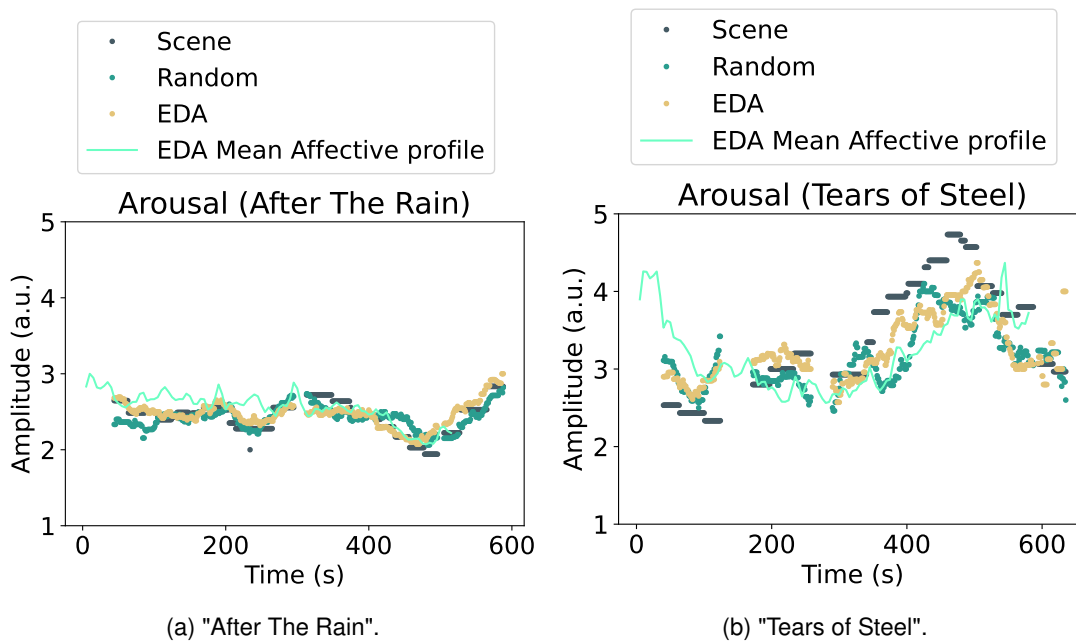


Figure 6.9: Examples comparing the collected MAP and the average arousal self-reports from the EmotiphAI Annotator.

Table 6.9: Comparison between the state of the art [326] and EmotiphAI Annotator best results on the MAP and arousal annotations correlation for the Pearson (Pearc. C) and Spearman (Spear. C). ToS (Tears Of Steel); AtR (After the Rain), ED (Elephant’s Dream); LA (LIRIS-ACCEDE); EAI (EmotiphAI).

Movie	Database	T (%)	Pearc.	Spear	Method
ToS	LA	66	0.55	0.64	Scene
	EAI	60	0.77	0.73	
AtR	LA	86	0.24	0.27	EDA
	EAI	50	0.80	0.65	
ED	LA	66	0.55	0.64	EDA
	EAI	90	0.26	0.25	

6.5 Discussion

The discussion of the results is divided into the evaluation of the EmotiphAI Annotator usability (RQ 3.1), accuracy tests (RQ 3.2), and the comparison between the content segmentation methods (RQ 3.3).

6.5.1 Usability Test

Addressing RQ 3.1, the results show that, although retrospective emotion annotation is a task that requires memory work and increases the subject mental workload, it is not overwhelming and too tiring for the volunteers as shown by the NASA-TLX of 40%. This value is in line with the state of the art, namely with EmoteU [19] (37.5% to 44.52%), and RCEA [298] (52.5% and 82.5%) and above [306], where live (31.6%) and textual (35.7%) annotation is performed. Amongst the two hypotheses under test (EDA and Scene-based segmentation), the EDA segmentation was the preferred method. Although the EDA method selected a higher number of clips for annotation than the Random method, it showed the lowest annotation time of all the methods. The SUS B+ score confirms the usability of the EmotiphAI Annotator for retrospective annotation, with the EDA method outperforming the Scene-based in terms of annotation time, SUS and preference score. The results demonstrate the usability of the EmotiphAI retrospective annotation tool for the annotation of long-duration content.

6.5.2 Validity Test

The RQ 3.2, regarding the EmotiphAI Annotator comparison to conventional approaches, was analyzed through a set of four tests:

Inter-subject Agreement

The evaluation error was lowest for the valence dimension indicating higher consensus. Overall, the evaluation error is lower than the reported in the state of the art (namely in [324]), although it should be noticed that different datasets and problems are addressed. The state of the art [328] reports that

in a naturalistic general content, slight disagreements in ratings can be expected, since many factors contribute to different ratings of the same stimuli, namely the subject's experience in emotion annotation, mood, personality, engagement and liking of the film, among others. The state-of-the-art reports low to average inter-rater agreement [139, 300, 328], increasing with the use of expert annotators [329], context [311], use of multi-modal information [311], rank annotation [297], data down-sampling [328], and outlier removal [130, 330]. After testing two alternative approaches, the authors in [306] report that for live annotation and textual annotation: *"The agreement of annotators remains very small, showing again the difficulty and inherent subjectivity of sentiment annotation"*. The findings within this chapter are in line with what is described in [328]: *"Experiments on several types of material provide information about their characteristics, particularly the ratings on which people tend to agree. Disagreement is not necessarily a problem in the technique. It may correctly show that people's impressions of emotion diverge more than commonly thought"*.

Self-report Coherence

The segmentation algorithms lead to similar annotations when averaged across the subjects' reports for a given timestamp (see Table 6.6). For each timestamp, low variability is observed, below a 0.2 amplitude on a [1, 5] scale. The high coherence can be explained since the same subjects annotate similar timestamps across algorithms, resulting in the annotation of roughly the entire content, with each timestamp being annotated by at least one subject, and in several cases by all (as seen in Table 6.4). The high annotation's coherence across the segmentation algorithms confirms the reliability of the retrospective annotation.

Comparison to the Reference Annotations

The STD between the EmotiphAI Annotator annotations and the state of the art (LIRIS-ACCEDE) (see Table 6.7) showed a maximum difference of around 0.65 in a [1, 5] scale. However, the value is still below the "critical" threshold (≈ 0.71) given by [324]. No meaningful difference was detected in the valence and arousal dimensions or across the different segmentation algorithms. The state of the art ground-truth was obtained in real-time and using continuous annotation [-1, 1], while the EmotiphAI Annotator uses a SAM discrete scale {1, 5} annotated retrospectively after the movie, which can introduce both latency and scale differences in the two annotations, and become a source of error in the observed STD between the two annotations. Nevertheless, overall, the experimental results are in line with the state of the art: *"Annotators agree on the trends but disagree on the values"* [324]; *"True emotion, however, does not necessarily fall into only one of the discrete emotion space sampling point"* [330].

Comparison to Electrodermal Activity

The EDA method outperformed the remaining, with the arousal self-reports showing lower-average to good correlation (0.26 to 0.8 correlation in [0, 1] scale) to the EDA data given by the MAP [326] in Table 6.8. The results reinforce the EDA as a marker of SNS activity given by the arousal dimension. The EmotiphAI Annotator Scene and EDA segmentation methods obtain competitive results comparatively with the state of the art, outperforming their results for two of the videos ("Tears of Steel" and "After the Rain"). For the "Elephant's Dream" video, the lower correlation score may result from the fact that the arousal average self-report values show minimum variability with no major emotional event (self-report values around the neutral state 3 for the entire content). The low variability annotations can mean that there was no emotion elicitation or low engagement in the video. The literature [326] reports that correlations to EDA are expected when an emotional event is detected by an increase or decrease of the emotion annotations from the neutral state (of magnitude 3 in EmotiphAI Annotator SAM scale). Overall, the EDA segmentation method can be of interest for affective computing applications, since it selects moments for emotion annotation correlated with changes in SNS activity.

6.5.3 Content Segmentation Method

Addressing RQ 3.3, amongst the segmentation methods, the EDA-based showed advantages, namely: 1) Lower annotation effort for the user as expressed by the lower annotation time comparatively to the remaining methods); 2) The obtained self-reports have displayed a higher correlation with the physiological dynamics, and the EDA data, serving as an indicator of SNS activity, is prone to selecting SNS-related events for emotion annotation; 3) Scene-based segmentation is limited on the use of a movie for emotion elicitation (unlike EDA-based); and 4) Allows to sort the events by intensity, which the literature has shown that are easier to recall [320], hence being important to show first to ensure that at least these are annotated (if the subjects do not comply with completing the rating of all the segments in a real-world annotation scenario).

It should be noticed that the EDA segmentation requires additional effort (and hardware) to record EDA data, which is not necessary for the other two methods. However, when collecting physiological data for the development of emotion recognition algorithms, such aspects can be mitigated. Overall, the annotation segmentation method should be chosen according to which better fits the research design and goals of the study.

6.6 Conclusion

Expanding upon the EmotiphAI Data Acquisition platform presented in the preceding chapter (Chapter 5), and the need for emotion recognition systems to have access to ground-truth annotations, this chapter focused on the annotation of emotional states. Moreover, it was identified a gap in annotation tools tailored for long-term naturalistic content, towards the annotation of naturalistic settings. In response, this chapter delves into *Obj 3. Emotion Annotation for Naturalistic Settings*, addressing RQ 3.1 to 3.3, which focuses on developing and evaluating the EmotiphAI Annotator, a novel tool for emotion annotation of long-duration content, in terms of its usability, accuracy, and comparative effectiveness against existing state-of-the-art solutions.

By focusing on long-duration content, rather than the short clips commonly used in current research, this method more closely mimics genuine emotional experiences. The EmotiphAI Annotator incorporates advanced content segmentation algorithms, which identify specific moments within the content for annotation. Moreover, the EmotiphAI Annotator employs a unique stepped retrospective approach, aiming to merge the benefits of real-time and retrospective annotation methods. This approach significantly reduces the overall number of annotations required, allowing for the annotation of longer videos by a broader range of individuals.

The platform was analyzed considering its usability and annotation accuracy. The experimental results showed that the EmotiphAI Annotator provides a good user experience with low mental workload and the retrospective annotations can be used as a reliable ground-truth estimate for emotion recognition systems. Among the segmentation methods, the EDA demonstrated low annotation effort and showed correlation to EDA data, ensuring that SNS-related events are selected for emotion annotation. As such, the EmotiphAI Annotator can enable quick, reliable and low mental workload emotion reports across longer elicitation content.

This platform could be the basis for widespread emotion annotation across diverse distributed real-world scenarios, leading to the creation of large databases for emotion recognition.

The gap in the literature for the sensor devices, group data collection, and annotation platforms of long-duration content, has resulted in the lack of public datasets with these characteristics. In the next chapter, a novel dataset for emotion recognition will be introduced, utilizing group physiological data and annotated long-duration videos collected in a naturalistic scenario.

Chapter 7

Group Emotion Dataset

Leveraging the platforms for data acquisition and annotation introduced in Chapters 5 and 6, this chapter tackles *Obj 4. Real-world Affective Computing Dataset*, specifically addressing RQ 4.1 and 4.2. To meet this objective, the G-REx dataset is presented, a dataset designed for the recognition of annotated group emotions using physiological signals in a natural setting.

The contents within this chapter were adapted, with permission, from:

- P. Bota, J. Brito, A. Fred, P. Cesar, and H. Silva, “A real-world dataset of group emotion experiences based on physiological data,” *Scientific Data*, vol. 11, no. 1, pp. 1–17, 2024

Contents

7.1	Introduction	132
7.2	Background	133
7.3	Methods	133
7.3.1	Dataset Design	134
7.3.2	Data Contents	139
7.3.3	Dataset Summary	143
7.3.4	Technical Validation	143
7.4	Results	146
7.4.1	Physiological Signals	146
7.4.2	Data Characterisation	148
7.4.3	Statistical Analysis	151
7.5	Discussion	152
7.6	Conclusion	156

7.1 Introduction

In recent years, the field of affective computing has gained prominence, namely in text sentiment analysis and image/video-based emotion recognition through body posture and facial expression. For this data, large data corpus are available (e.g. AffectNet [331] with 0.4 million annotated facial expressions; EmotiW challenge [332] with 1088 annotated videos), which is crucial for the development of accurate artificial intelligent algorithms.

This growth has also been observed in emotion recognition based on physiological data, namely in unobtrusive physiological sensors (e.g. EDA or PPG) which aim to capture data in real-life settings. In the physiological-based affective computing literature, as detailed in Section 2.3 there is a large number of publicly available datasets (e.g. [131, 137, 140, 333, 334, 335, 336] in Section 2.3). However, the majority of these datasets are designed for data collected in the lab and rely on the use of short clips/images validated to elicit basic emotions, e.g. CASE [131], Emognition [333], or POPANE [334]. As denoted in Chapter 6, the use of short video clips does not replicate a naturalistic emotion elicitation setup, where normally there is a build-up of emotion. Moreover, although in-lab experiments allow higher control over the collected data, the literature has questioned whether they can be replicated and generalized to real-life scenarios [6, 140].

The state-of-the-art has been moving from in-lab controlled and small-video excerpts setups to alternative data collection closer to naturalistic scenarios, such as the BIRAFFE2 [136] (Chapter 7), in which physiological data was collected (ECG, EDA) during video games, the PPB-Emo dataset [335] (Chapter 7) with physiological data collected (EEG) during driving, or K-EmoCon [137], with (EDA, PPG, ECG, skin temperature, EEG) data collected during a social debate. A limited number of robust datasets with data collected in real-life scenarios can be found, such as the DAPPER [140].

This part of the thesis aims to address this gap by introducing a labelled naturalistic dataset specifically designed for group emotion recognition through physiological data – the G-REx dataset, guided by the exploration of the following research question:

RQ 4.1: *Can large amounts of annotated physiological data be reliably collected in a naturalistic setting using the EmotiphAI platform?*

RQ 4.2: *How does an infrastructure designed for group physiological data acquisition perform in a real-world setting?*

In line with the naturalistic data collection, the experiment relies on real movies for emotion elicitation, using physiological sensors integrated into an unobtrusive and wireless bracelet device (EmotiphAI Wearable Chapter 5), and collecting data in a group setting (analogous to a cinema theatre). Then, the data is annotated retrospectively by the volunteers, which allows an undisturbed visualisation of the movie and the annotation of only selected movie scenes. This paradigm allows for a naturalistic emotion

elicitation over a long period (each movie has around 2 hours). The dataset contains data from over 190 subjects, covering 31 movie sessions and more than 380 hours of physiological data. The proposed experimental setup for annotated affective data collection can be replicated across diverse naturalistic experimental settings, from a cinema session to a classroom or a hospital.

7.2 Background

The terms "in-the-wild"/"real-world" or "naturalistic" data have been denoted to describe data collection when the experimenters do not control the emotion elicitation nor constraint the data acquisition [6]. These can be further divided into "ambulatory" settings where the data is collected in daily living with the subjects moving freely, or "static" when the data collection is limited to a specific location (such as workplace, car or cinema) [6]. The authors in [337], compared stress responses induced in the lab to stress induced at the volunteer's home. The experimental results showed that the volunteer in-lab HR during the stressor was lower than in the real-world setting. Similarly, in [338] the authors observed that physiological data collected in the lab differed from data collected replicating real-world data, where the person is free to move. This resulted in the model created in the lab deteriorating its emotion classification performance when tested with real-world-like data.

Another example of the limitations of in-lab data collection setups is the [339] MOSAIC program, which had the goal of evaluating affect detection systems in real-world scenarios. The data was collected following a lifelog setup, during everyday routines, such as working and work-related travelling. Emotion-related ground truth was collected for positive and negative affect, stress and anxiety. These were measured at the start of the study and once per day using the experiment sampling technique. Neither of the teams met the program goal metrics for affect detection, attaining accuracy near zero across teams. Based on the results, the authors in [339] suggest the need for a different data collection paradigm that is closer to a real-world scenario.

7.3 Methods

The G-REx dataset was designed to bridge the gap between the real world and controlled experiments. This is done by opting for collecting data in long-duration content in a cinema, retrospectively annotated in small segments by their emotional relevance. Moreover, annotated EDA and PPG data is collected unobtrusively as detailed in Chapters 5 and 6.

7.3.1 Dataset Design

The G-REx dataset was designed for large data collection in naturalistic group scenarios. As a proof-of-concept, data was collected in a room on a university campus replicating a movie theatre (Figure 7.1). Instead of relying on selected movie clips as observed in the literature [305, 333, 334], longer duration content is used, in particular, a movie.

The dataset was collected to expand the annotated public datasets for physiological-based affective computing collected in naturalistic scenarios and propose an experimental setup that can be easily replicated across any naturalistic setting with little effort to the user.

Data Annotation

Datasets on emotion recognition usually contain emotion-annotated segments. As previously described in Chapter 6, the common practice described in the literature for short clips is to perform real-time continuous annotation of arousal and valence (e.g. [131, 305, 333, 134]). However, to perform the annotation in naturalistic scenarios or as in the case of a long film, the annotation becomes tiring and distracts the user from the elicitation process.

To address this issue, the platform proposed in Chapter 6 is used. In this approach, the annotation is performed by the subjects themselves and is as quick and not intrusive as possible. Instead of annotating the entire two-hour movie, which would be too time-consuming and tiresome for the volunteer, the subject annotated only selected segments of a few seconds (20 seconds) where events (onset events with high amplitude) were detected on the subjects EDA data (segment of [onset - 5 seconds, onset + 15 seconds]). A time interval of 20 seconds was selected to replicate the annotation duration in AMIGOS [87], and taking the 5 seconds previous to the onset from the literature reporting the latency period of emotional stimuli between 1 to 5 seconds [318].



(a) G-REx data collection cinema room.



(b) Placement of the EmotiphAI Wearable.

Figure 7.1: Photos taken during the G-REx data collection movie sessions.

Participants Recruitment

The data collection was performed between October 2022 and June 2023, making a total of 31 sessions. The volunteers were recruited from participants in the Diferencial¹ cinema sessions. The Diferencial is a student club from the IST of the University of Lisbon. The cinema sessions were advertised on Diferencial's social media platforms, namely Twitter and Instagram, being free and open for anyone to participate, regardless of whether they are students at the university. In the post description, there was a notification that physiological data collection was being performed on volunteers. At the cinema sessions, a lab researcher of the team approached the audience, described the experiment and asked if they would like to participate. At the end, a chocolate bar was given as a reward for the participation.

Ethics Statement

The study was submitted to and approved by the IST - University of Lisbon Ethics Committee (Ref. n.º 11/2021 (CE-IST) Date: 20/04/2021). The participants were given a consent form upon arriving at the cinema room before the data collection. The informed consent form contained information regarding the context, goal, procedure, data registration, data privacy, and risks of participation in the experiment. In this form, the participants manually filled in their participation agreement, age, gender, if they participated in the study with any friends and their familiarity with the movie. Additionally, the participants were asked to fill out a physiological data purpose form with an agreement for the usage, visualization and analysis of the data, sharing the data in academic publications, at conferences, in media, and with external partners. Participants were notified that their participation was voluntary, had to be done in an informed and free manner, and that their data could be destroyed and their participation withdrawn by request at any time without consequences.

Data Collection Setup

The data collection took place on an amphitheatre at IST. The cinema sessions were performed once per week starting at 8 to 8.30 PM during the school academic year. To collect and annotate the data, an adapted version of the EmotiphAI platform described in (see Chapter 5) was used. Moreover, the platform was adapted to the ScientISST device. The EmotiphAI [12] platform was set up in a corner of the room, with the router at the centre. The hardware and software used in the experiment are detailed next.

¹<https://diferencial.tecnico.ulisboa.pt>; Accessed on 20/02/2024

Hardware

- **EmotiphAI Collector** (Microsoft Surface Pro, 7 1.10GHz x 8 CPU): was used to host the EmotiphAI platform for both physiological data collection [12] and emotion self-reporting (Figure 7.2 – A). A local network was created using a TL-WR940N router (Figure 7.2 – B), allowing the EmotiphAI Wearable to communicate with the EmotiphAI Collector.



Figure 7.2: EmotiphAI platform setup used in the Differential cinema sessions. Nomenclature: EmotiphAI Collector (A); TL-WR940N router (B); EmotiphAI Wearable based on the ScientISST device (C).

- **EmotiphAI Wearable** (Figure 7.2 – C): contained two physiological sensors that were connected to the subject's skin using Pre-gelled Ag/AgCl electrodes (to improve the skin conductivity and electrode's adherence to the skin) on the non-dominant hand with the respective sensors. The EDA sensor electrodes were placed on the thenar and hypothenar areas, and the PPG was placed surrounding the index finger distal phalange. The details of each sensor are described in Table 7.1. For this data collection, the R-IoT device described in Chapter 5 was no longer available, and for that reason, it was replaced by ScientISST CORE (Figure 7.2 – C), a device designed to collect physiological data². The ScientISST device provided a few advantages against the R-IoT

²<https://scientisst.com>; Accessed on 20/02/2024

device, namely, it allows both Transmission Control Protocol (TCP) (to control for data loss) and Bluetooth communication and is an open-source device. Moreover, the ScientISST device allows the integration of over 6 input channels at a high sampling rate (reported sampling rate > 10000 Hz for one device).

Table 7.1: Overview of EmotiphAI Wearable characteristics. Nomenclature: Sampling rate (SR); Resolution (Res.); Communication (Comm.).

Characteristic	Sensor	Range	Bandwidth / Specs	Additional Info
SR: 100 Hz	EDA	0 – 25 μ S	0 – 2.8 Hz	Input Voltage Range: 1.8 – 5.5 V
	PPG	0.3 – Vdd	Input Voltage Range: 3 – 5.5 V	
Res.: 12 bits Comm.: WiFi (TCP), Range: 100 meters Size: 22.51 x 40.90 x 53.50 mm Weight: 64g				Battery: Li-On; 7.4 V 800 mA

Software

- **EmotiphAI Collector:** The physiological data collection was controlled by the EmotiphAI Data Acquisition platform [12] (Chapter 5) adapted for the ScientISST device. The data was stored locally on the device and then moved to a private cloud. For further detail on the acquisition platform, the reader is referred to Chapter 5 [12].
- **EmotiphAI Annotator:** The annotation platform runs on the collected data in post-processing. After the movie is over, the annotation platform iterates across the users and, for each, their most significant moments as given by high amplitude EDA events (onset to peak amplitude) are selected for the volunteers to annotate retrospectively. For further detail on the annotation platform and segment selection methodology, the reader is referred to Chapter 6 [15].

Experimental Protocol

Figure 7.3 shows the experimental procedure for physiological and emotional data collection at each cinema session. A research team of two to three members was on the site to help follow each step of the protocol.

1. **Consent Form:** As each participant arrived for the cinema session, they were approached by one member of the research team, given a description of the experiment, and asked if they would like to participate. If they agreed, an informed consent form was given for the volunteers to read and fill out.

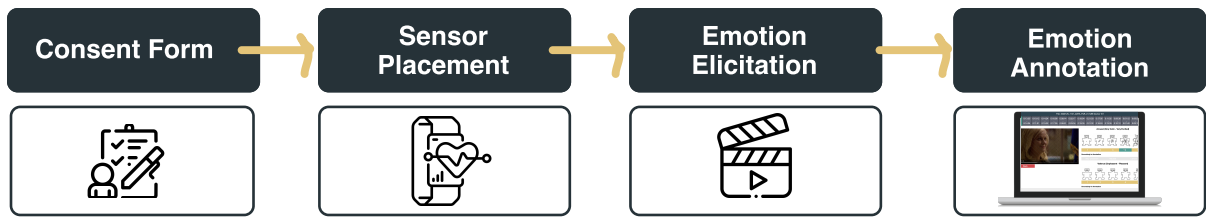


Figure 7.3: Experimental protocol for the collection of the G-REx dataset.

2. **Sensor Placement:** Upon agreeing to the data collection, the EmotiphAI Wearable was placed on the non-dominant hand with two physiological sensors (EDA and PPG). The volunteers were then ready to start watching the movie and start the data collection.
3. **Data Collection:** The movie and physiological data collection were manually synchronized by starting both at the same time. The volunteers watched the movie in a naturalistic scenario, being able to sit in any location in the amphitheatre, surrounded by their friends or strangers.
4. **Data Annotation:** At the end of the session, the volunteers were approached by our team to request the annotation of emotional segments in terms of arousal, valence and uncertainty in the annotation. A description of the emotion annotation is displayed in Table 7.2.

Table 7.2: Description of G-REx self-reported emotion annotation scales.

Category	Description	Range
Arousal	Denotes general energy deactivation/activation [340]	$[1, 5] \in \mathbb{N}$
Valence	Denotes displeasure/pleasure [340]	$[1, 5] \in \mathbb{N}$
Uncertainty	Denotes the level of uncertainty/certain in the emotion annotation	Yes/No

5. **Follow Up:** The literature on emotion recognition has shown that the personality type might influence the emotional reaction (e.g. neuroticism is correlated to high negative-emotional response) [341, 342]. With this in mind, on the day after the data collection, a follow-up email was sent to the volunteers. The follow-up email had the goal of thanking the participants for their contribution, and sharing an optional questionnaire with the big-five factor model for personality assessment using the 50-item English version of the International Personality Item Pool (IPIP-J) [343].

Overall, each session lasted approximately 2 hours and 30 minutes, which included around 10 minutes for emotional self-reporting, and an additional 10 minutes for sensor placement and the completion of the consent form.

Affective Stimuli

The cinema movies were selected by the Diferencial team. Most of the movies were part of thematic cycles, each dedicated to specific themes. The collected data covered the following cycles: "Horror",

"Ghibli", "Mind*uck", "Musical", "Asian culture", and "Is Anybody out there?". Additionally, two collaborations were performed with IST student groups ("AmbientallST", "NucleAr" and "AEIST") and one collaboration with a production company ("JumpCut"). In two sessions, a short film was displayed before the movie data collection.

7.3.2 Data Contents

The G-REx dataset was organized to contain both the raw and transformed data. As well as all the code for the data transformation. To guarantee the reproducibility and transparency of the research, the code utilized for the data transformation is also included within the dataset, with the resultant plots and table content derived during the research process.

The G-REx dataset is available on *Zenodo*³. The data will be made available after completing the End User License Agreement (EULA) available⁴. For the structure of the dataset, a similar approach to the one in [134] was followed, and an overview is shown in Figure 7.4. The dataset contains six main folders, outlined as follows:

- **1_Stimuli:**

- *Raw/video_info.json/csv* – Contains detailed information on the movies used in the dataset.
- *Transformed/stimu_trans_data_<DATA_TYPE>.pickle* – Contains the information of the movie details for the annotated segments and session data. $\langle DATA_TYPE \rangle \in \{session, segments\}$.

- **2_Questionnaire:**

- *Raw/quest_raw_data.json/csv/xlsx* – Contains the questionnaire data for all the participants in the dataset.
- *Transformed/quest_trans_data_<DATA_TYPE>.pickle* – Contains the user ID and device information of the emotion-annotated segments and session data. $\langle DATA_TYPE \rangle \in \{session, segments\}$.

- **3_Physio:**

- *Raw/S<X>_physio_raw_data_M<Y>.hdf5*; where $\langle X \rangle$ is the session ID $\in [0, \dots, 28]$, and $\langle Y \rangle$ is the movie ID $\in [0, \dots, 30]$ – Contains the raw HDF5 data collected by the EmotiphAI platform for each session X and movie Y .
- *Transformed/physio_trans_data_<DATA_TYPE>.pickle* – Contains the transformed raw and filtered EDA, PPG, HR and time information data for the annotated segments and session data. $\langle DATA_TYPE \rangle \in \{session, segments\}$.

³<https://zenodo.org/record/8136135>

⁴<https://forms.gle/RmMosk31zvvQRaUH7>; Accessed on 20/02/2024

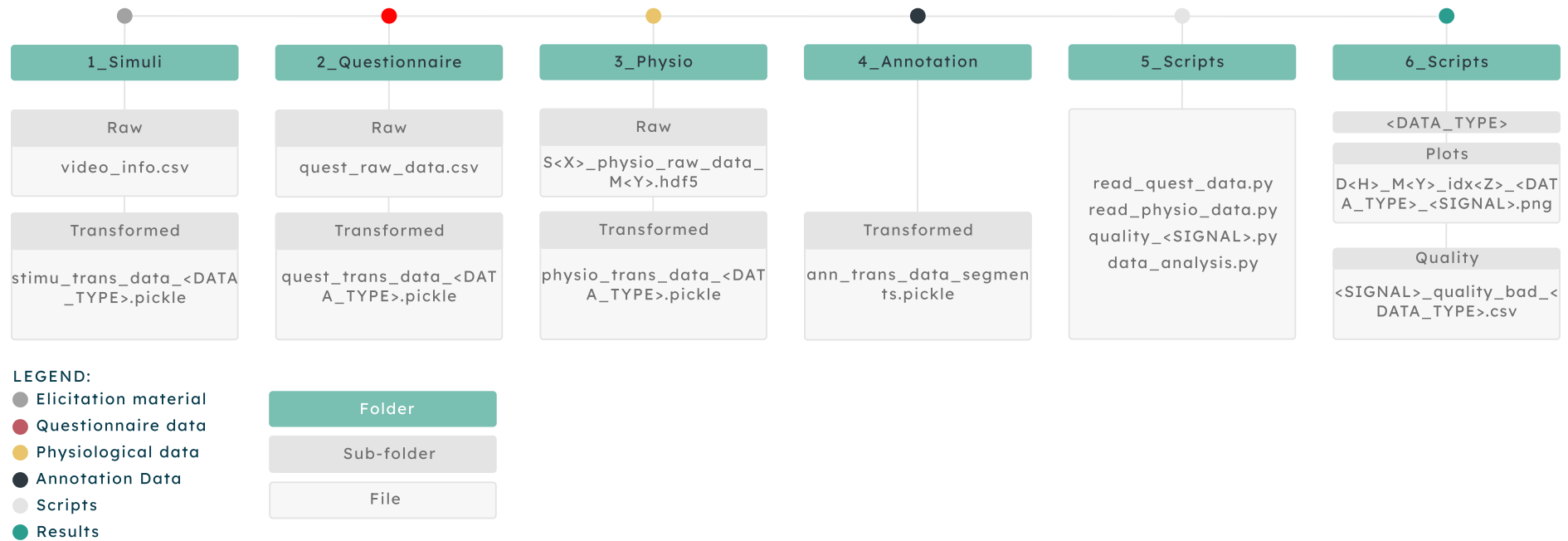


Figure 7.4: G-REx data structure.

- **4_Annotation:**

- *ann_trans_data_segments.pickle* – Contains the arousal, valence and uncertainty values for the annotated segments.

- **5_Scripts:**

- *read_quest_data.py* – Returns the video and user questionnaire information.
- *read_physio_data.py* – Reads the collected raw physiological signals, emotion annotations, video data and the self-report questionnaires to store the transformed data in separate dictionaries, with matrices data for the session and annotated data segments.
- *data_analysis.py* – Script used to obtain the plots and tables displayed in the technical validation section.
- *quality_<SIGNAL>.py* – Obtain the data quality and lower quality signals. $\langle \text{SIGNAL} \rangle \in \{\text{EDA}, \text{PPG}\}$.

- **6_Results:**

- *<SIGNAL>/<DATA_TYPE>/Plots/D<H>_M<Y>_idx<Z>_<DATA_TYPE>_<SIGNAL>.png* – Plot of the raw and filtered data. The $\langle H \rangle$ variable is the device ID; $\langle Y \rangle$ is the movie name; and $\langle Z \rangle$ is the sample index; $\langle \text{SIGNAL} \rangle \in \{\text{EDA}, \text{PPG}\}$; $\langle \text{DATA_TYPE} \rangle \in \{\text{session}, \text{segments}\}$.
- *<SIGNAL>/<DATA_TYPE>/Quality/<SIGNAL>_quality_bad_<DATA_TYPE>.csv* – Table with the physiological signals technical validation results. $\langle \text{SIGNAL} \rangle \in \{\text{EDA}, \text{PPG}\}$; $\langle \text{DATA_TYPE} \rangle \in \{\text{session}, \text{segments}\}$.

Each pickle file consists of a dictionary. The detailed information regarding the keywords of each pickle file can be seen in Figure 7.5.

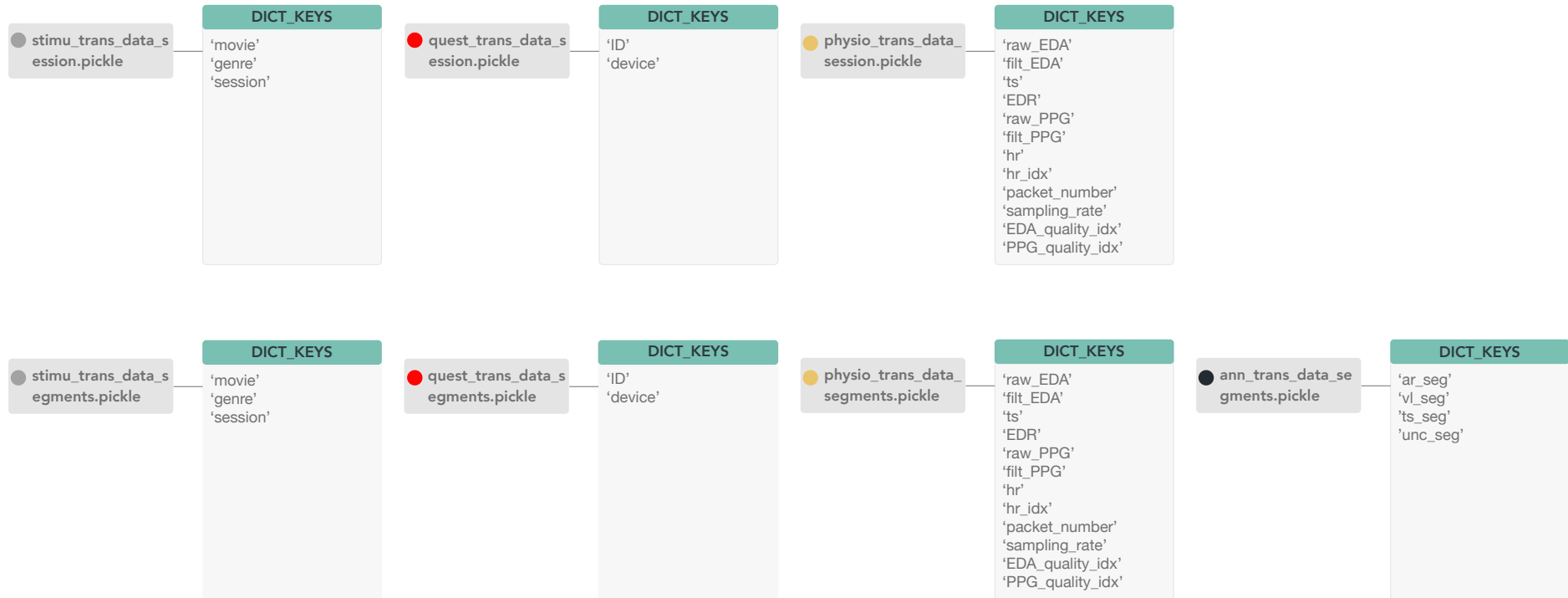


Figure 7.5: G-REx dataset pickle data structure with respective keys.

7.3.3 Dataset Summary

The G-REx dataset consists of data collected from 31 sessions, making a total of 191 users and more than 380 hours of data collected. This includes data from the physiological signals EDA and PPG, and emotional annotations of selected movie moments, making a total of 1400+ annotated segments. In two of the sessions, two different movies were seen, while in the remainder only one movie was seen resulting in a total of 31 sessions but 29 movies. A summary of the total data collected can be seen in Table 7.3.

Table 7.3: Summary of the G-REx dataset data.

	Total	After Processing		
	Session	Segments	Session	Segments
# of Participants	191	191	92	149
# of Movies	31	31	27	27
# Sessions	29 (\approx 384.5 hours)	29 (\approx 8.2 hours)	25 (\approx 175.9 hours)	25 (\approx 5.7 hours)
# Samples	241	1481	112	1031
Age Range			18 – 69	
Physiological Signals			EDA, PPG	
Emotion Annotations			Arousal, Valence	

Preprocessing

The raw data was collected using the EmotiphAI Collector and stored on a *HDF5* file [344] (Chapter 5). To facilitate the usage of the data, the latter is stored in a compact format of the session and segments data, as well as questionnaire, stimuli and emotion annotations on separate but synchronized dictionaries stored on *pickle* files.

The raw data in *xlsx* or *HDF5* format was transformed into dictionaries containing the relevant data in matrices stored in a *pickle* format.

For the physiological data pre-processing, the *BioSPPy* library [8] was used for filtering the EDA and PPG signals, peak extraction and EDA decomposition into the EDR and EDL components. The analysis of statistical tests was conducted using the *SciPy* library [345]. The processing was done in Python 3.7.4, and the required code is available in the *5_Scripts* folder so it can be easily replicated.

7.3.4 Technical Validation

To characterize the collected data, physiological data was categorised into sets of higher and lower quality. Furthermore, description metrics are utilized to compare the two sets. The distribution of emotion annotations and movie genres is then observed, highlighting the diversity and comprehensiveness of the dataset. Finally, statistical evaluations were conducted, comparing the distributional properties of

arousal, valence, and genre groups, benchmarked against mean EDA, HR, and subjectively reported arousal and valence measures.

Physiological Signals

The EDA signal is read through the connection of gel electrodes to the skin, making the signals easily subjected to artefacts such as loss of contact of the electrodes with the skin either due to a long-duration use or sweat, which disconnects the gel and adhesive. Similarly, the PPG signal can be affected by the tightness of the bracelet, which can cause saturation at the maximum value, or be loose and lose connection to the finger. In addition, the PPG allows the extraction of the HR, which has expected range-values denoted in the literature between 40 and 200 beats per minute (bpm) [346]. Values outside this range are noise and can be removed.

To remove noisy data and obtain a view of its description, quality metrics identified in the literature [12, 346, 347, 348, 349, 350, 351] were applied for the EDA and PPG signals. Using the quality metrics, the data was divided into quality and lower-quality sets. As observed in the literature, some works apply cut-offs to remove low-quality data using a cut-off threshold [348], while others [352] use the metrics to obtain an overview of the data distribution of the two to observe how the quality segments and lower quality segments distributions compare to each other.

Additionally, the technical validation of physiological signals encompasses two distinct data collection formats: session data and 20-second annotated segments. Following Nasser et al. [348], a threshold of four seconds was adopted to identify lower-quality data within the 20-second segments. However, given that session data spans approximately two hours, a four-second criterion was considered negligible. Consequently, a more conservative threshold of 7% was applied to assess the session data quality, ensuring relevance and accuracy in a longer-duration analysis.

Statistical Analysis

Taking arousal, valence and movie genre scores as groups with possibly different EDA and HR physiological measures, a statistical test for a normal distribution of the different groups was computed using the Shapiro test⁶, to test the null hypothesis that the data was drawn from a normal distribution. When any of the obtained p-values were below the thresholds for significance level (i.e. $p < 0.05$) the null hypothesis was rejected, and it was concluded the data is not normally distributed. Taken that the data is not normally distributed, the Kruskal-Wallis H-test⁷ was obtained to assess the null hypothesis that the population median of all of the groups is equal. On the other hand, when the p-value is above 0.05,

⁵https://github.com/PatriciaBota/physio_group_emotion_phd/tree/main/group_emotion_dataset; Accessed on 20/02/2024

⁶<https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.shapiro.html>; Accessed on 20/02/2024

⁷<https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kruskal.html>; Accessed on 20/02/2024

Table 7.4: Data quality metrics deployed for analysing the EDA and PPG signal quality in the G-REx dataset. The code used to obtain the quality metrics is available in the *5_Scripts* folder, or the chapter's repository⁵.

Signal Cut-Off	Metric	Description
EDA & PPG	Full scale	Amplitude at the bit resolution of 12 bits ($2^{12} = 4095-1$) for more than 4 seconds and 7% in the session data [347]. Applied to the filtered data.
EDA & PPG	Zero	Amplitude below $0.05 \mu\text{S}$ for EDA and 0.01 a.u. for the PPG for more than 4 seconds in the segment data and 7% in the session data. Applied to the raw data.
EDA & PPG	Loss	Counting packet number and time discontinuities [12]. Data was considered as lower quality if it had a data loss above 7%.
PPG	Abnormal heart rate (HR)	HR below 40 bpm or above 200 bpm detected for more than 4 seconds in the segments or 7% in the session data was considered as low-quality data [346].
Data Distribution		
EDA & PPG	Signal-to-noise Ratio (SNR)	For the EDA, SNR was obtained by the logarithmic ratio between the cleaned signal (low-pass 4 th order Butterworth filter of 5 Hz and 0.75 seconds-window smoother moving average) [8] and noise (3 rd order Butterworth band-pass filter of 2-10 Hz) [350]. For the PPG, the noise was obtained by a 4 th order high-pass Butterworth filter with 15 cut-off frequency [353] and the cleaned signal by a 4 th order band-pass Butterworth filter with 1 – 8 cut-off frequency [8].
EDA & PPG	Max, Mean, Min	Statistical features (maximum, mean and minimum) extracted from the filtered data [347].
PPG	Spectral Entropy	Entropy of power spectrum between 0.1 to 3 Hz, ranging between 0 for a periodic signal and 1 for a constant spectrum [348, 349, 351].

not showing enough evidence to reject the null hypothesis that the data was drawn from a normal distribution, the Analysis of Variance (ANOVA) test⁸ was performed. All the tests were computed using the annotated segments data.

7.4 Results

The technical validation of the G-REx dataset is presented in this section. The results are divided into two main sections: the physiological signals, data characterisation, and a brief statistical analysis of the physiological signals and emotion annotations.

7.4.1 Physiological Signals

The physiological signals are analysed in terms of the data quality metrics and their distributional properties.

Electrodermal Activity

The experimental results for the data quality analysis on the sessions and emotion segments can be seen in Table 7.5 for the EDA data.

1. **SNR:** The experimental findings reveal that the SNR for the EDA signal stands at approximately 120 dB, indicating that the signal's intensity surpasses the noise level by around 120 dB in both the session and annotated segment data. The SNR is lower for the lower-quality data, registering about 100 dB for the session data and further declining to roughly 40 dB for the segment data, both exhibiting a greater standard deviation in the lower-quality data. This contrasts with the SNR range of approximately 50 to 60 dB reported by Behnke *et al.* [334] and the average SNR span of 26.66 dB to 37.74 dB across all signals documented by the authors of Emognition [333], aligning with or even falling below the values reported for the higher-quality data. Similarly, research by Gautam *et al.* [350] identified an SNR range from 50 to 80 dB, while Castro-García *et al.* [352] reported an SNR of 61.6 dB for EDA data. However, it should be noted that the SNR values can vary depending on the specific methodology employed to calculate the SNR metric.
2. **Full scale:** Across all the data no saturation was detected. Saturation is commonly observed when the range of the sensor values is below the reading, for example by a hard press on the sensor. The results show a correct positioning of the sensor, with data in line with the physiologically expected.
3. **Zero:** The zero percentage allows the detection of records where no data was collected. The percentage of zero-valued signals is low, attaining a value below 1% for the quality data in both

⁸https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.f_oneway.html; Accessed on 20/02/2024

the sessions and annotated segments data. On the other hand, a high percentage of zero data (%) is observed on the samples identified as lower-quality data (above 50%). This shows that the zero-data metric is the most contributing factor for the classification of lower-quality data. The sensor's zero reading may result from several factors: the electrodes losing contact with the skin, physiological conditions like hyperhidrosis leading to electrode disconnection due to excessive sweating during extended periods (e.g., during a 2-hour movie), or issues associated with the device's form factor, such as a being loose or a broken cable. Additionally, it should be taken into consideration that due to the naturalistic setting of the data collection, some of the devices lost WiFi connection (e.g., volunteers leaving the room), and there were cases where the device was not on any participant or had no battery, leading to no data being recorded.

The literature [354] expects a prevalence of hyperhidrosis at 4.8%, following a study in the United States of America (USA) population in 2016. However, this value might be biased towards a severe case, with 70% reporting severe excessive sweating. In this study, an unexpectedly high percentage of hyperhidrosis incidents may stem from the device's reliance on gel electrodes for skin connectivity, with the duration of a two-hour movie session potentially exceeding the electrodes' capacity to maintain contact. Alternatively, these occurrences could be attributed to specific sensor or device characteristics, such as the low resolution or gain of the ADC.

4. **Loss:** Across the dataset, a data loss of 0% was observed, a value significantly lower than those reported in the literature. For instance, Bottcher *et al.* [349] documented an average data loss of around 50% for data streaming. In alignment with our findings, Castro-García *et al.* [352] also reported a loss of 0%.
5. **Max, Mean, Min:** The lower quality data in both the sessions and annotated segments show a lower average value (around 0 μ S). Similarly, the maximum and minimum values are higher in the quality data when compared to the lower-quality data.

Photoplethysmography

A similar analysis was performed for the PPG data. The quality metrics can be seen in Table 7.4, and the obtained results in Table 7.6.

1. **SNR:** The difference in the SNR in the noisier and quality data sets are predominant, with the quality data showing a SNR above 70 dB and the lower quality data SNR below 10 dB.
2. **Full scale:** Once again no saturation is identified, similarly to [347], where a very low full-scale percentage was observed.

3. **Zero:** A higher number of zeros in the lower-quality data in both the sessions and the annotated segments was observed, while the quality data has a lower percentage of zeros (below 2%). This can be explained by, in some of the sessions, devices were turned on but were not being used by any volunteer, thus recording a value of 0 throughout the entire session.
4. **Abnormal HR:** Abnormal values for the HR (defined to be below 40 and above 200 bpm [346]) are not observed in the quality data, while for the lower-quality data, abnormal values are observed for almost the entire set (around 90%).
5. **Spectral Entropy:** Higher entropy is observed for the quality data. The authors in [348], denote that a lower entropy is expected for the quality data, highlighting a pointier spectrum in its amplitude waveform. While a flat spectrum (i.e. uniform distribution) is characterized by higher entropy and noisier data. This was not observed in the data. The higher entropy observed can be a result of the increased complexity in the quality data, which shows higher standard deviation and diverse morphology compared to a flat line (lower quality data). Our results are in line with Bottcher *et al.* [349], which reports a threshold of around > 0.8 to denote a lower-quality signal.
6. **Max, Min, Mean:** The data shows lower amplitude values for the lower data quality set, attaining a minimum value of zero.

7.4.2 Data Characterisation

After the pre-processing step of the noisy sample removal, a total of 1031 annotated segments and 112 sample sessions were obtained. A histogram with the total number of annotated samples can be seen in Figure 7.6a. The figure shows that, overall, the users annotated around 5 to 10 samples per session. The number was tailored to 7 following the volunteer's feedback on their preferences. With a few exceptions of users who participated in more than one session. For example, one subject who was part of the cinema club participated in most of the sessions, which is seen by the large peak in the histogram near ID 25. The IDs in "blank" correspond to users who were removed on the data pre-processing step or did not annotate any segments.

Each movie was assigned to its predominant genre following the *IMDB* characterization⁹. In Figure 7.6b, the assigned movie genre of each annotated sample is shown. As observed, data from eight main movie genres were collected due to the movie cycles.

Figure 7.7a shows the volunteer's personality scores across the big five dimensions. The extraversion dimension shows a broader range of values covering most of the scale, followed by neuroticism. These dimensions have been correlated to the expression of emotions, namely to the frequency and intensity

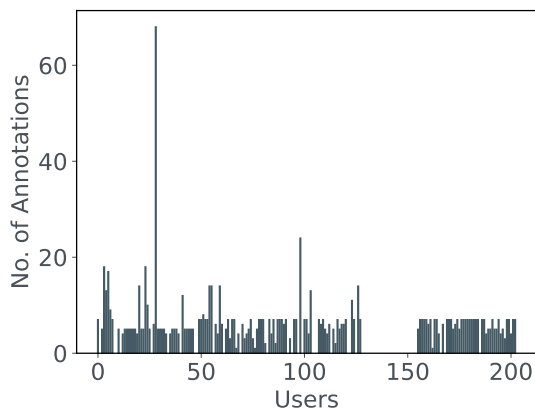
⁹<https://imdb.com>; Accessed on 20/02/2024

Table 7.5: EDA data quality metrics analysed in the G-REx dataset.

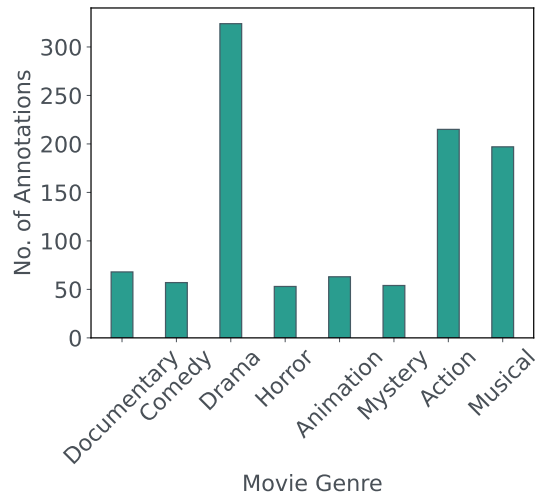
Data	Size	SNR (db)	Full scale (%)	Zero (%)	Loss (%)	Max (μS)	Mean (μS)	Min (μS)
Quality Segments	1265	122.71 \pm 23.26	0.0 \pm 0.0	0.3 \pm 1.84	0.0 \pm 0.0	7.85 \pm 5.22	6.42 \pm 4.25	5.48 \pm 3.89
Lower Quality Segments	216	37.79 \pm 35.3	0.0 \pm 0.0	83.46 \pm 22.4	0.0 \pm 0.0	2.94 \pm 5.3	0.7 \pm 1.71	0.0 \pm 0.01
Quality Session	136	127.12 \pm 15.09	0.0 \pm 0.0	0.75 \pm 1.7	0.0 \pm 0.0	15.54 \pm 6.59	7.06 \pm 4.3	2.03 \pm 3.41
Lower Quality Session	105	100.56 \pm 39.71	0.0 \pm 0.0	52.06 \pm 32.3	0.0 \pm 0.0	9.97 \pm 7.51	1.75 \pm 1.96	0.0 \pm 0.0

Table 7.6: PPG data quality metrics analysed in the G-REx dataset.

Data	Size	SNR (dB)	Full scale (%)	Zero (%)	Loss (%)	Abnormal HR (%)	Spectral Entropy	Max (a.u.)	Mean (a.u)	Min (a.u.)
Quality Segments	1213	76.07 \pm 14.24	0.01 \pm 0.17	1.62 \pm 3.16	0.0 \pm 0.0	0.0 \pm 0.0	0.68 \pm 0.09	1903.95 \pm 1137.23	24.09 \pm 55.37	-1244.47 \pm 745.14
Lower Quality Segments	268	3.6 \pm 14.39	0.03 \pm 0.32	95.57 \pm 17.61	0.0 \pm 0.0	93.66 \pm 24.37	0.05 \pm 0.18	196.01 \pm 809.31	1.65 \pm 15.63	-150.52 \pm 607.29
Quality Session	199	72.47 \pm 9.92	0.01 \pm 0.08	1.21 \pm 1.55	0.0 \pm 0.0	0.0 \pm 0.0	0.72 \pm 0.07	3518.95 \pm 1122.89	14.93 \pm 27.25	-2539.92 \pm 774.41
Lower Quality Session	42	7.92 \pm 20.47	0.01 \pm 0.05	89.37 \pm 28.96	0.0 \pm 0.0	85.71 \pm 34.99	0.10 \pm 0.25	605.43 \pm 1543.48	1.72 \pm 6.01	-432.49 \pm 1104.39



(a) User ID

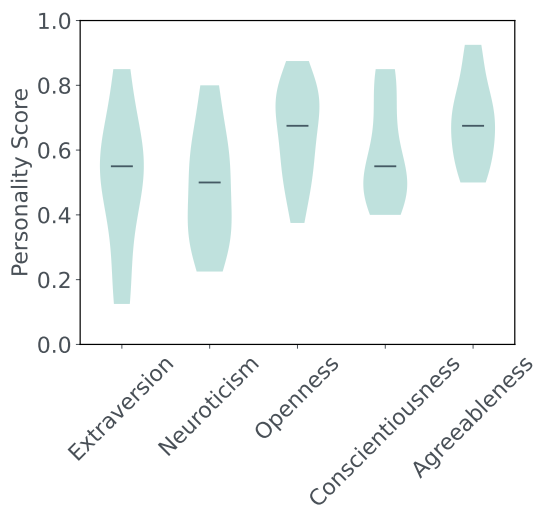


(b) Movie genre

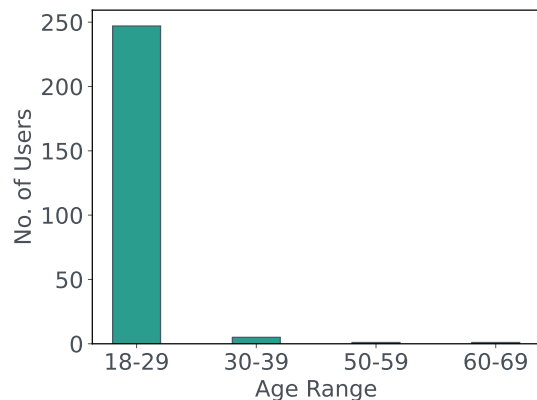
Figure 7.6: Histogram with the number of annotations per user and samples collected per movie genre.

of positive and negative emotions [341, 355]. The remaining dimensions are skewed to the upper range of the scale.

Lastly, Figure 7.7b shows the age range of the volunteers. As can be seen, data was acquired across all age distributions, with 18 to 29 years old being predominant. This range is expected since the data collection took place on a university campus.



(a) Personality



(b) Age

Figure 7.7: Distribution of personalities and age range.

Emotion Annotation

After the movie, the EmotiphAI Annotator [15] was used to annotate selected movie scenes using the SAM scale [356] on the arousal and valence dimensions. The results of the annotations can be seen in

Figure 7.8, where both dimensions cover the entire annotation space. Statistical characterisation of the annotations distribution is shown in Table 7.7.

As can be seen, both dimensions are centred around the value of 3, with a standard deviation of around 1. Both dimensions show negative (left-modal) skewness and kurtosis. The valence shows a near-zero skewness corresponding to a symmetrical distribution, while arousal is slightly negatively skewed. Regarding the kurtosis score, both dimensions show an elevated negative kurtosis score. A high negative value describes a flatter distribution compared to a normal distribution, denoting a more homogenous distribution of the data across the annotated scale with the probability of values near the mean lower than in a normal distribution. Moreover, the distribution has lighter tails, suggesting fewer extreme values. Gatti *et al.* [305], also analyze the annotations of kurtosis and skewness, obtaining an overall negative skewness and a positive kurtosis distribution for valence and arousal. However, it should be taken into consideration that these metrics are heavily impacted by the number of collected samples and the elicitation content.

Table 7.7: Statistical metrics extracted from the G-REx annotations.

	Mean	STD	Skewness	Kurtosis
Arousal	3.13	1.15	-0.29	-0.68
Valence	3.30	1.06	-0.08	-0.57

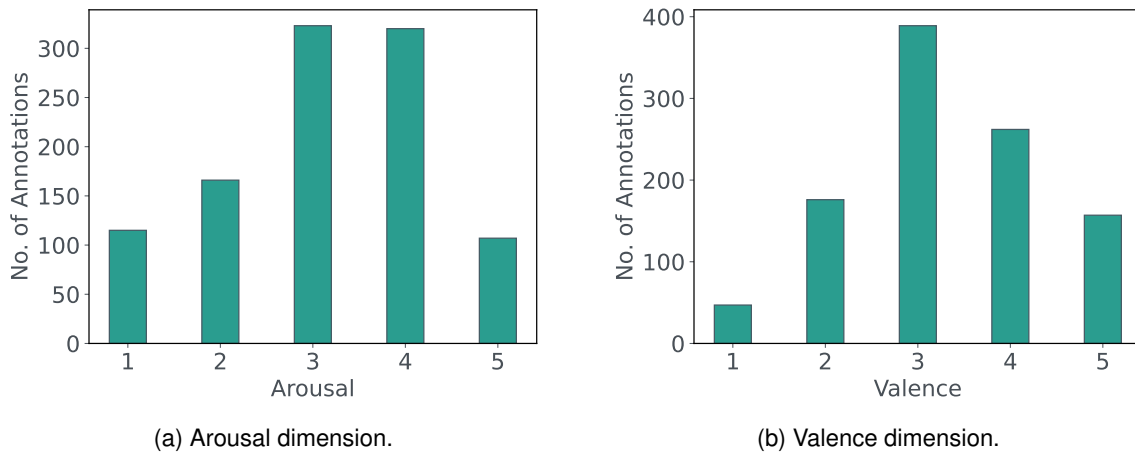


Figure 7.8: Histogram of the self-reported G-REx annotations.

7.4.3 Statistical Analysis

The experimental results for the statistical test on the normal distribution and group annotation differences (arousal, valence and movie genre distribution for physiological measures) are shown in Table 7.8. Analyzing the Shapiro test, across most groups, a p-value was obtained below the 0.05 threshold, denoting that the groups do not follow a normal distribution. These results are in line with the normality

test performed in [305], where a Kolmogorov-Smirnov normality test p-value below the threshold was obtained across participants.

For the group differences (Kruskal-Wallis/ANOVA), it was observed that, overall, the null hypothesis of equal medians across groups could not be rejected ($p\text{-value} > 0.05$). The normalized mean EDA and HR were found to have very similar medians across the different groups, namely arousal and valence scores from 1 to 5. These results are expected since emotion classification is a complex task and a more diverse set of features, combined with artificial intelligence algorithms are required to separate the different classes and perform emotion recognition. An exception is the normalized mean HR and EDA, and the arousal and valence self-reports for the different movie genres where a p-value below the threshold was obtained. Denoting that at least one group's population median is different from the others. To better understand the statistical results from Table 7.8, the data distributions of the groups across the studied measurements are illustrated in Figures 7.9 to 7.12.

Table 7.8: Statistical test p-values determining whether there are statistically significant differences among the different groups. The tested groups are arousal and valence dimensions $\in \{1, 2, 3, 4, 5\}$; Movie genre: Drama, Animation, Musical, Comedy, Horror, Mystery, Action, Documentary. The most significant results are shown in bold ($p\text{-value} < 0.05$). A Kruskal-Wallis test is performed when the normality test (Shapiro) obtains at least for one group a p-value < 0.05 , and the ANOVA if all the p-values > 0.05 . Tests were computed using the annotated segment data.

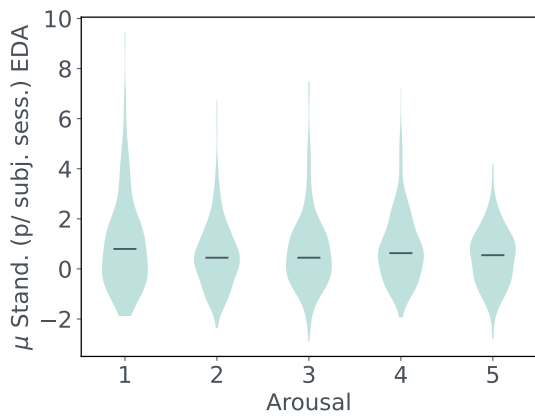
Group	Measurement	Shapiro	Kruskal-Wallis ANOVA
Arousal	Mean EDA	(0.00, 0.00, 0.00, 0.00, 0.87)	0.17
	Mean HR	(0.00, 0.00, 0.00, 0.00, 0.00)	0.22
Valence	Mean EDA	(0.00, 0.00, 0.00, 0.00, 0.00)	0.15
	Mean HR	(0.01, 0.00, 0.00, 0.00, 0.00)	0.55
Genre	Mean EDA	(0.23, 0.51, 0.01, 0.80, 0.00, 0.00, 0.02, 0.0)	0.00
	Mean HR	(0.22, 0.19, 0.08, 0.00, 0.00, 0.00, 0.07, 0.00)	0.00
	Arousal	(0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00)	0.00
	Valence	(0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00)	0.00

7.5 Discussion

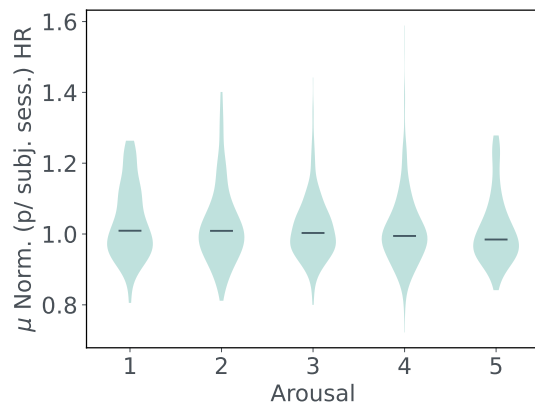
A detailed description of the videos used in the *video_info.csv* file was made available, so they can be identified by the users.

Physiological Data Evaluation

Addressing RQ 4.1 (on whether data can be collected reliably in a naturalistic setting using the EmotiphAI platform), it can be seen that the EDA data is in line with what is reported in the literature. Namely, Braithwaite *et al.* [357] reports expected EDA values between 2 to 20 μS , increasing for periods of high arousal. In [358], the authors report EDA data between 10 and around 28 μS (M (Mean): 15.55; STD:

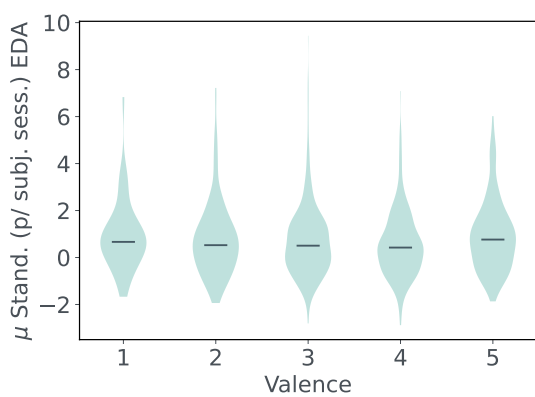


(a) Mean EDA – Arousal

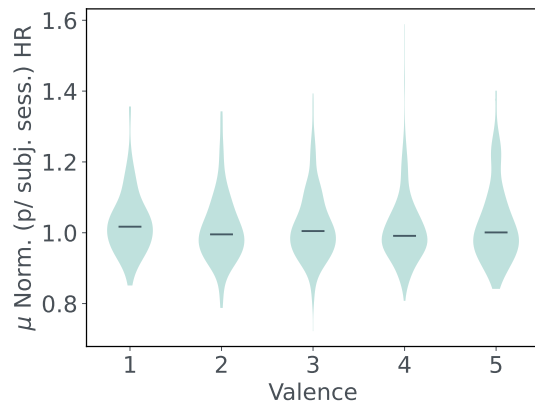


(b) Mean HR – Arousal

Figure 7.9: The data distribution for the mean standardized EDA—achieved by subtracting the session mean and dividing by the session standard deviation for each subject—and normalized HR—obtained by dividing by the session mean for each subject—was analyzed across arousal self-report scores. This analysis presents the group medians and their extremal values, with the assessment conducted using data from annotated segments.

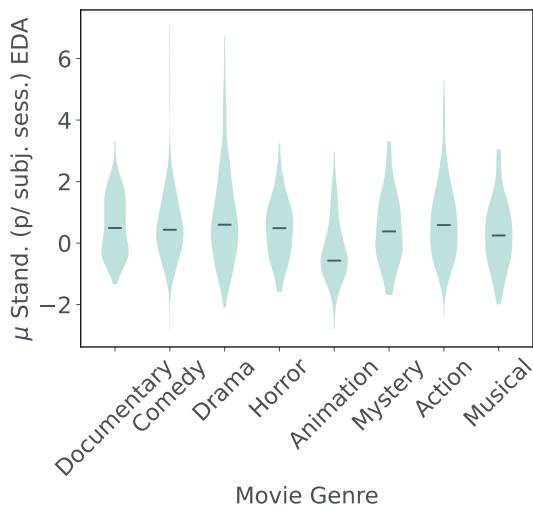


(a) Mean EDA – Valence

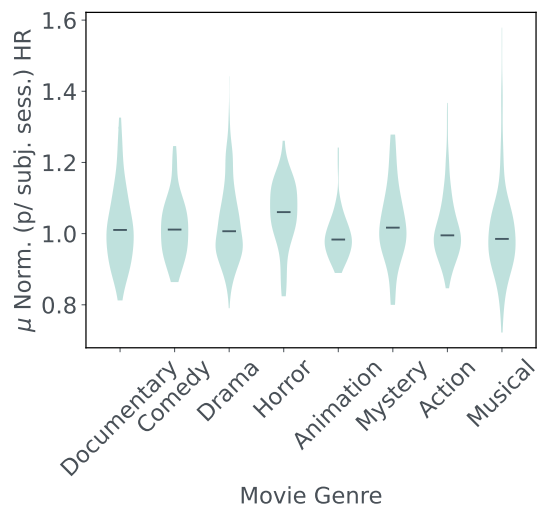


(b) Mean HR – Valence

Figure 7.10: Data distribution for the standardized mean EDA and normalized HR across the valence self-reported scores.

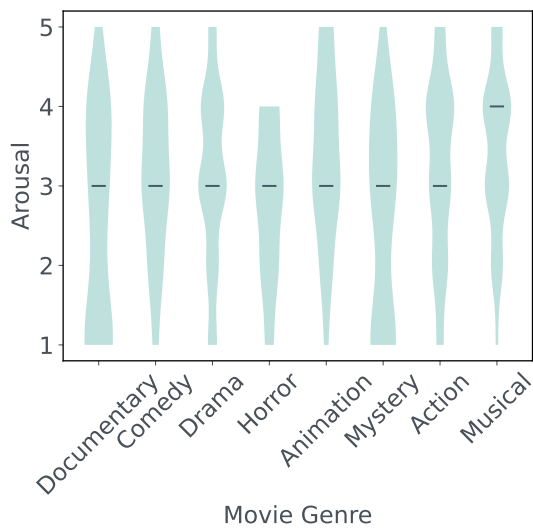


(a) Movie Genre – Mean EDA

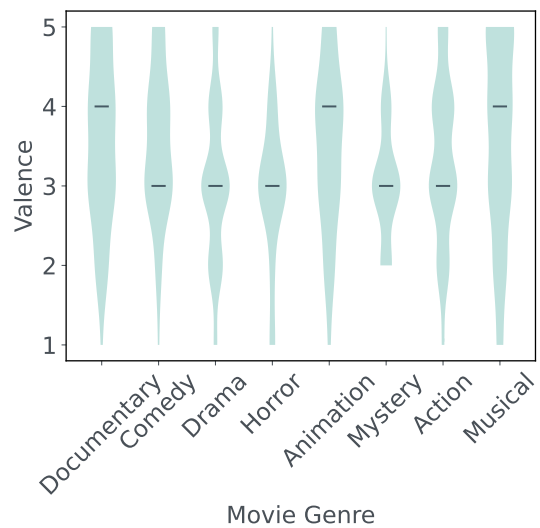


(b) Movie Genre – Mean HR

Figure 7.11: Data distribution for standardized mean EDA data per subject and normalized mean HR per subject across the movie genres.



(a) Movie Genre – Arousal



(b) Movie Genre – Valence

Figure 7.12: Data distribution for the arousal and valence self-reports across the movie genres.

1.67). These values are in line with the values obtained for the quality data, taking that deviations can result from the type of electrodes or their body placement.

Overall, the annotated segments from the EDA sensor exhibit a smaller proportion of lower-quality segments in comparison to the session data. This is primarily because data exhibiting zero-values for more than 7% of the session duration were excluded from the session set, based on a minimal threshold criterion established to identify low-quality data. Throughout a two-hour data collection session, it is expected for subjects to experience periods of relaxation or boredom, leading to episodic drops in EDA to zero.

For the segment data, the threshold was set as four seconds in 20 seconds of data. These values were selected following the literature [348]. The authors in [348, 349] report around 64% to 75% of quality EDA data. In the current work, a value of around 85% was obtained for the segments and around 55% for the session data. While for EDA data in the sessions set, large amounts of data were discarded due to the zero-metric, the same was not observed for the PPG data. This led to a lower number of sessions and annotated segments being identified as lower quality for the PPG data. For the PPG data, the authors in [349] report around 50% of quality data using an Empatica⁵ device, while the proposed dataset obtained around 80% for the segments and session PPG data.

The different parts of the data collection protocol and their required software/hardware are susceptible to specific constraining factors, which it is detailed below. The dataset limitations are related to collecting a dataset in a naturalistic scenario.

Data Collection Setup

Taking into consideration that the data collection was performed in a naturalistic setting, the volunteers were free to cover themselves and the devices with their clothes such as jackets and even leave their seats. Moreover, the movie was played by the cinema club on a separate projector. So it was necessary to start the movie and the data collection with a manual cue that directed the simultaneous start. This may present issues in the precision of timing that should be noted. However, these are not a major concern given the timescales of the physiological responses measured. Nonetheless, future work should focus on using an automatic method for synchronization.

Emotion Annotation

Emotion annotation was performed retrospectively based on the subject emotional events expressed by their EDA. Taking into consideration the naturalistic setting of the data collection, some of the detected events may not be related to emotional events but to random movements. Additionally, when performing the emotion annotation the volunteers were given a video preview of the selected moments for annota-

tion. However, as the volunteers were freely viewing the video, their emotion elicitation could have been not from the movie but from conversations with their peers. To gather insight on these issues, in the emotion annotation platform, an open text box is provided where the participants can introduce long textual external comments for each video segment, and the information on which participants participated in the experiment sat side by side is annotated.

Participants Health and Room Conditions

This part of the thesis aimed at devising a data collection methodology that can be replicated at a large scale in the real world. As such, in the proof-of-concept, the volunteers participated in a cinema session where the goal was to collect as little data as possible. This approach ensured minimal interruption to the normal cinema experience and encouraged a large number of volunteers to participate regularly in the data collection. Consequently, no records were obtained about any psychiatric or neurological conditions of the participants, nor was there any information gathered on whether participants had consumed any pharmacological medication during or prior to the study. Similarly, no data was acquired on the humidity, temperature of the room, or food intake during or prior to the movie. Such factors can potentially influence physiological responses and, thereby, emotional status and future works should contemplate incorporating this information if the protocol setup allows reducing this limitation while maintaining the efficiency of the data collection process.

7.6 Conclusion

This chapter consolidates the advancements introduced in the preceding chapters, showcasing the deployment of the EmotiphAI Data Acquisition platform (Chapter 5) and the EmotiphAI Annotator (Chapter 6). Building upon these platforms, *Obj 4. Real-world Affective Computing Dataset*, namely RQ 4.1 and 4.2 are addressed by the creation of a dataset of physiological data within a naturalistic group setting named the G-REx dataset.

The dataset contains data collected from over 190 subjects across 31 movie sessions, culminating in more than 380 hours of physiological data. To analyse the data, the data was divided into sessions and annotated segments, validated through quality signal metrics. The results showed that the data was collected reliably, with a low percentage of zero data and no data loss. The physiological data was in line with the expected physiological values, and the annotations were distributed across the entire scale.

The G-REx dataset paves the way for varied and complex emotion recognition analyses using physiological data, much closer to the nuances of real-world situations than previously available. The G-REx dataset demonstrates the feasibility of collecting large amounts of data in naturalistic settings, which is crucial for applying advanced deep learning techniques that require large datasets to train. The G-REx

affective data has the potential to be used across diverse domains, from the entertainment industry, to assess the emotional impact of movies, video games, and other media as long as it is recorded, offering insights into the audience's emotional responses, to the development of personalized content recommendation systems that better resonate with the viewers', enhancing engagement and immersion. Furthermore, the group context of the data collection can aid the emotion recognition performance, by introducing a holistic view of the subject emotional state, as well as the emotional state of the group as a whole. Lastly, the group's physiological synchrony can be explored to identify meaningful events or moments within the content, such as in movies or live performances.

On the whole, the G-REx dataset not only moves forward the field of affective computing but also contributes to a broader understanding of human emotional responses in collective environments, offering invaluable insights for a wide range of applications where emotional engagement and interaction are key.

Chapter 8

Conclusions and Future Work

This chapter presents the conclusions of this thesis along with the explored RQs, outlines potential future work and applications of the developed technologies, and discusses the main take-home and challenges that lay ahead for further research in the field of affective computing.

Contents

8.1	Revisiting the Research Questions	159
8.2	Discussion	166
8.2.1	Input Data	166
8.2.2	Emotion Annotation	166
8.2.3	Emotion Classification	167
8.2.4	Moving Forward	167
8.2.5	Affective Computing Applications	168
8.3	Future Work	169

8.1 Revisiting the Research Questions

The primary goal of this thesis was to explore group emotion recognition based on physiological data. This was performed over four main objectives, addressed over 8 chapters, across 12 RQs.

In Chapter 3, this thesis embarked on **Obj 1. Development of Affective Computing Algorithms**, by exploring the state of the art in affective computing. this exploration was performed around the following RQs:

RQ 1.1: *What feature set/machine learning algorithm should be applied for emotion recognition?*

The RQ 1.1 was tackled by first performing a systematic review of the state of the art in affective computing, where the most predominant datasets, algorithms, sensor modalities, and evaluation metrics,

among others, were identified. The results indicated that the publicly available datasets are the ITMDER, WESAD, DEAP, MAHNOB, and the Eight-Emotion EESD. The most explored algorithms are the SVM, K-NN, and random forest, and predominant sensors are the EDA, ECG, and respiration. Upon this survey, the following RQ sought to uncover:

RQ 1.2: *What performance can be achieved by predominant datasets in the literature?*

Addressing both RQ 1.2, a large set of features (570 EDA, 373 PPG, 322 ECG, and 487 respiration) were extracted from the physiological data. These were further reduced through a feature selection step (sequential forward feature selection) to perform arousal and valence binary classification. The experimental results showed that most features are data-dependent, being selected, on the whole, once per dataset. This impacted the performance of the classifiers on the different datasets, with similar results obtained across classifiers and sensors.

Additionally, across all modalities, the valence dimension attained the highest prediction scores. For the WESAD dataset, the arousal dimension's F1-score dropped to 0.0 compared to the accuracy value, indicating that the class labels were largely imbalanced. For the ITMDER dataset, the proposed methodology was able to surpass the state of the art or obtain very competitive results for both the arousal and valence dimensions. The good results in the ITMDER and WESAD valence dimensions, with the F1-score surpassing the accuracy score, indicated that the class labels are imbalanced and are biased in the results. For the EESD, DEAP, MAHNOB and WESAD, the current work provided the first results for the selected sensors using the datasets in the literature.

The sensor modalities were then combined to explore a multi-modality classification:

RQ 1.3: *What is the best method to deal with multi-modal data for emotion classification?*

In addressing RQ 1.3, the findings highlighted the effectiveness of multi-modality over single-modality configurations, with performance being contingent on the specific dataset and emotion dimension being considered. This observation aligned with existing literature, suggesting a consensus on the benefits of multi-modal approaches [89]. Specifically, for the ITMDER and DEAP datasets, both decision fusion and feature fusion techniques demonstrated superior performance compared to state-of-the-art results, with valence dimension classifications showing particularly good outcomes. For the MAHNOB dataset, the current work was not able to surpass the results of the literature, however, competitive results were obtained. No direct comparisons could be drawn for the EESD and WESAD datasets due to a lack of similar studies. However, the comparison between single and multi-modality approaches underscored the potential of the latter to enhance or maintain classification results across various datasets and dimensions. The use of feature fusion not only matched or exceeded the performance of decision fusion but also offered significant advantages in computational efficiency, with average execution times notably lower, suggesting a more practical approach to be applied in real-world applications.

Chapter 4 further explored the **Obj 1. Development of Affective Computing Algorithms**, by considering the impact of group environment in emotion recognition. This work delved into the concept of group emotion recognition, by analysing how to measure physiological synchrony, analysing diverse similarity metrics and data representations to measure physiological synchrony for emotion recognition. This was performed through the following RQ:

RQ 1.4: *What synchronization metrics and data representations are most suitable for measuring physiological synchrony for emotion recognition?*

The RQ 1.4 was addressed by exploring the WGS methodology. This approach determines the subject's emotional label by considering the emotional states of group members and weighing them based on their physiological synchrony with the unknown subject. To perform this analysis the AMIGOS and K-EmoCon datasets were selected as they were the identified public datasets with group physiological data available in the literature. From the sensor modalities explored in RQ 1.1 to 1.3, the HRV and EDA features were selected for further analyses. The HRV can be extracted from both PPG or ECG (AMIGOS collected ECG data, K-EmoCon PPG's data), and along with EDA can be extracted in single-point configurations, in non-obtrusive places like the hand or wrist. Thus, facilitating data continuous data collection in naturalistic scenarios, as it is the goal of this thesis. Moreover, to meet the data imbalance identified in the previous RQs, the M-F1 score was used as the primary metric for evaluating the classification performance, which equally weights the precision and recall of each class.

The experimental results on the AMIGOS dataset with data from groups of four subjects visualizing a movie excerpt (from 14 to 23 minutes long), showed that learned representations based on HRV features emerged as most effective for arousal, while a combination of EDA and HRV features was superior for valence. Notably, the strong performance of EDA for valence contradicts conventional expectations of its association with arousal, suggesting a potential influence of high arousal-valence annotation correlation or lower variability in data on these findings. For the K-EmoCon dataset, which contained data from a dyadic debate (around 10 minutes), average pooling returned a comparable performance to WGS using cosine similarity on HRV features for arousal, whereas cross-correlation in the EDL space was most effective for valence.

The study concluded that feature space representations generally outperformed morphological or image-based representations, with Pearson, Spearman, and cosine similarity metrics often reducing performance. Non-weighted approaches to group synchronization like average pooling showed comparable to superior results, the latter for the case of the poor-performance synchronization metrics. Consistently, valence dimensions achieved higher classification scores than arousal, aligning with literature expectations.

Based on the learned representations from RQ 1.4, the WGS approach (integrating group context) was compared to the traditional intrasubject methodology (with no group information).

RQ 1.5: *Does the emotion classification accuracy improve with the inclusion of group-level information?*

Addressing RQ 1.5, the WGS methodology demonstrates an enhanced performance or competes closely with existing benchmarks. Specifically, it outperforms reported accuracies in the literature, such as Gupta *et al.* [287] findings, by relying solely on group emotion labels without additional multi-modal data. This is particularly notable in the AMIGOS dataset, where, to the writer's knowledge, it is the first study to exclusively use group data for emotion recognition, marking a significant advancement over traditional intrapersonal models that overlook group dynamics.

The interpersonal model consistently surpassed intrapersonal models across both datasets and most dimensions, except for the valence dimension in the K-EmoCon dataset, where data of dyadic conversation is used. For instance, in the AMIGOS dataset, interpersonal models achieved approximately 72.15% and 81.16% for arousal and valence, respectively, compared to the intrapersonal models' 59.40% and 66.44%. Similarly, in the K-EmoCon dataset, the interpersonal approach yielded improvements in arousal (52.63% vs. 47.67%) but observed a slight decrease in valence performance (65.09% vs. 73.70%) compared to intrapersonal models. This highlights the potential of leveraging group-level data for enhancing emotion recognition accuracy, suggesting a promising direction for future research in affective computing.

Taking that these methods require data to be created, a survey of the literature was performed to identify existent datasets and devices for group physiological data collection. The survey identified a gap, with very few datasets focusing on group physiological data (only AMIGOS with groups of four, K-EmoCon with dyadic groups were identified), this was a reflection of the lack of devices for group physiological data collection of EDA and HR-related data.

This led to **Obj 2. Group-based Physiological Data Collection**, explored in chapter 5. To meet the gap identified in the literature, the EmotiphAI platform was created. EmotiphAI is a small and mobile, low-cost infrastructure for group physiological data collection and real-time visualisation, integrating the EmotiphAI Wearable, EmotiphAI Collector, and EmotiphAI User Interface. Moreover, the EmotiphAI was developed to be easily deployed in naturalistic settings. The EmotiphAI platform was evaluated through the following RQs:

RQ 2.1: *How does the sampling period affect the data loss and transfer quality in multi-device scenarios?*

Addressing RQ 2.1, data was collected for 1, 10, and 20 devices at different sampling rates, namely 200 Hz, 60 Hz, and 25 Hz, to assess the impact of the sampling period on data collection. The results indicated that the sampling period significantly affects the data loss and transfer quality in multi-device scenarios. Moreover, the EmotiphAI platform was able to reliably collect data with low data loss and

uniform data transmission in both individual and group settings with 1 device up to 200 Hz (5 ms). When the number of devices was increased to 10 and the sampling rate was maintained at 200 Hz, the data loss increased to around 65%. Requiring that for 10 devices, the sampling rate should be decreased to 60 Hz to maintain a low data loss (below 1.5%). When the number of devices was increased to 20, only the 25 Hz sampling rate maintained a low data loss (below 0.5%).

These results surpassed the state of the art, namely the device FMCI by Xinhuanet which only allowed the collection of data from 1 sensor (EDA) at 1 Hz up to 20 devices, and BITalino (6 analogue sensors to which any sensor can be connected), with around 4 devices at 100 Hz.

Next, it was explored whether the network infrastructure through which data is transmitted affects the performance of the EmotiphAI platform (RQ 2.2):

RQ 2.2: *To what extent does the network infrastructure influence the maximum number of devices that can collect data without significant data loss?*

The experiments revealed that a single-antenna router performs comparably to a three-antenna router when handling one device at a 200 Hz sampling rate, with both maintaining data loss below 1%. Likewise for 10 devices, where the 1-antenna router can maintain a high-quality transmission at 60 Hz. However, this level of performance does not extend to scenarios with 20 devices, where the single-antenna router is only reliable for a sampling rate of 10 Hz. On the other hand, the introduction of a repeater router for 10 devices, allows a boost in the sampling rate from 60 Hz to 100 Hz, while keeping data loss below 1%. This was not observed for 20 devices, where even with a bridge router, data collection remained unstable for sampling rates faster than 25 Hz.

Having developed and validated the infrastructure for group physiological data collection and the requirement that emotion recognition algorithms require annotated data, this thesis moved to **Obj 3. Emotion Annotation for Naturalistic Settings:**. In Chapter 6 the EmotiphAI Annotator was developed to address this objective. The EmotiphAI annotator is a web-based tool focused on annotating content in naturalistic settings. Current state-of-the-art annotation tools usually perform single post-hoc annotation or live annotation of excerpt clips. Both of these methods have limitations, as they either lose the dynamic context of the emotion or are too distracting and exhausting for the subject, not being fit for a longer-term naturalistic stimulus, like a movie. The EmotiphAI Annotator was developed to address this gap, by allowing the retrospective annotation of selected moments of the content, thus, simplifying the annotation process and reducing the annotation effort. Three different content segmentation methods were explored to assess which moments are more suitable for emotion annotation, namely: random segmentation, Scene-based (temporal segmentation), and EDA-based (physiological segmentation). The EmotiphAI Annotator was evaluated through the following RQs:

RQ 3.1: *Are retrospective annotations in long-duration content usable for emotion annotation?*

To address RQ 3.1, the EmotiphAI usability was evaluated through the subject mental workload using the NASA-TLX questionnaire. The results indicated that while retrospective emotion annotation is a task that requires memory work and increases the subject mental workload, it is not overwhelming and too tiring for the volunteers, as shown by the NASA-TLX of around 45%. This value is in line with the state of the art, namely with EmoteU [19] (37.5% to 44.52%), and RCEA [298] (52.5% and 82.5%) and above [306], where live (31.6%) and textual (35.7%) annotation is performed. Moreover, the usability of the tool was assessed through the SUS questionnaire, which resulted in a B+ score, indicating that the tool is effective, efficient and satisfactory for its users.

Next, the reliability of the EmotiphAI Annotator's annotations was assessed:

RQ 3.2: *How do retrospective annotations compare to conventional approaches in long-duration content?*

The reliability of the EmotiphAI platform was analysed through a set of four metrics: inter-subject agreement, self-report coherence, comparison to the reference annotations, and comparison to EDA.

Starting by the inter-subject agreement, following the work by Grimm and Kroschell [324], this metric was measured by the annotation's STD in a given timestamp, and the evaluation error given by the difference between the average standard deviation in a timestamp and the optimal standard deviation bound. The results indicated that the evaluation error for both dimensions is low (less than half the distances between two discrete SAM numbers), and notably low for the valence dimension, indicating that it can be more reliable than the arousal dimension. Moreover, the evaluation error was lower than existing state-of-the-art benchmarks [324], with the caveat that comparisons are somewhat limited due to differences in datasets and problem contexts.

The self-report coherence among the annotations for different content segmentation methods was also measured through the STD in a given timestamp and the evaluation error by comparing the average of the user's annotations for the three algorithms, in each data timestamp. The results indicated that the segmentation algorithms result in similar annotations with low variability observed. This coherence is attributed to subjects annotating similar content across algorithms, ensuring comprehensive coverage of the material.

The comparison to the reference annotations was performed by comparing the annotations of the subjects to the reference annotations given by the LIRIS-ACCEDE benchmark, where the data was collected for the same elicitation content, once again through the STD in a given timestamp, and the evaluation error. The results showed that the annotations from the EmotiphAI Annotator show a maximum standard deviation of 0.65 on a 1 to 5 scale when compared to the LIRIS-ACCEDE benchmark, which is below the critical threshold of discrepancy. No significant differences were found in valence and arousal or between segmentation algorithms, despite methodological differences that could introduce variance. These results are in line, following the observed in the coherence evaluation.

Finally, the comparison to the EDA was performed by comparing the annotations to the MAP, as a metric of SNS activity which is related to emotional arousal. The EmotiphAI Annotator's Scene and EDA segmentation methods deliver competitive results, particularly for videos that elicit strong emotional responses. Comparing the MAP to the self-reports, the EmotiphAI Annotator obtains a higher correlation with values among lower-average to good correlation (0.26 to 0.8 correlation in [0, 1] scale), surpassing the correlation values for the annotations collected in the work by Ting *et al.* [326] for two of the movies ("Tears of Steel" and "After the Rain").

Taking the reliability of the annotations, the content segmentation methods were compared to understand which is more suitable for emotion annotation in long-duration content, through the following RQ:

RQ 3.3: *Which content segmentation method is more suitable for emotion annotation in long-duration content?*

Among the analysed segmentation methods, i.e. Random, Scene and EDA-based, the latter stood out by requiring less annotation time, demonstrating a higher correlation with physiological responses, and possibly in identifying emotionally significant events for annotation. Although it necessitates additional effort and hardware to record EDA data, this investment is justified when the objective is to develop emotion recognition algorithms and physiological data is being collected. The choice of segmentation method ultimately hinges on the specific requirements and objectives of the research.

Building upon the developed EmotiphAI platform, namely the Data Acquisition and the EmotiphAI Annotator, the thesis moved to **Obj 4. Real-world Affective Computing Dataset** in Chapter 7, where the G-REx dataset was created. The G-REx dataset is a novel dataset that contains physiological data collected from groups in naturalistic settings, annotated retrospectively. The dataset was evaluated through the following RQ:

RQ 4.1: *Can large amounts of annotated physiological data be collected reliably in a naturalistic setting using the EmotiphAI platform?*

RQ 4.2: *How does an infrastructure designed for group physiological data acquisition perform in a real-world setting?*

Addressing RQs 4.1 and 4.2, the G-REx dataset was collected in a naturalistic setting, with the subjects watching movies in a University cinema session. The dataset comprises data from 190 subjects across 31 movie sessions, totalling over 380 hours of physiological recordings. The data was annotated using the EmotiphAI Annotator on the user's own devices, retrospectively. The collected data quality was assessed through diverse metrics from data loss to signal quality (e.g. full-scale and zero-values, SNR, and abnormal values), with the quality data obtained values in line with the state of the art ([346, 347, 348, 349, 350, 351]).

This proof of concept showcased the EmotiphAI platform's capability for collecting annotated physiological data in naturalistic experimental setups. Thus, paving the way for advancing the field of group emotion recognition through physiological data that closely replicates real-world experiences for the development of more accurate and impactful affective computing systems.

8.2 Discussion

This thesis addresses the field of emotion recognition, from the theoretical foundations and infrastructural developments, to the creation of annotation and classification methodologies for group emotion through physiological data. This comprehensive approach underscored the multifaceted nature of the emotion recognition field and provided an in-depth understanding of the current challenges and opportunities.

8.2.1 Input Data

The nuanced nature of emotions demands a multifaceted approach to data collection. Emotion theories, such as psychological constructionism [55, 75], posit that emotional responses are not only physiological but are intricately linked to behaviour and the environmental context. This thesis underscores the importance of capturing a comprehensive view of emotions by integrating group data as environment information. This approach aligns with the understanding that emotions are adaptive responses to situational cues, suggesting that a richer dataset that incorporates further information on body behaviour and its environment could lead to a more meaningful view of emotion and improve emotion classification performance.

8.2.2 Emotion Annotation

The task of annotating physiological data with emotional states presents notable challenges. The arousal-valence model, while foundational, is often difficult for participants to understand due to the abstract nature of these dimensions. This confusion is compounded by individuals' biases towards their self-perception, the decay of emotional intensity over time, and the disruptive nature of real-time annotation. Furthermore, the distinction between emotions and moods is critical; emotions are brief, event-specific experiences that deplete a physiological response, whereas moods are longer-lasting and not tied to specific events and/or physiological reactions. In long-term data collection, there is a tendency to annotate moods rather than emotions, leading to a mismatch between the physiological signals captured and the annotations provided. This discrepancy suggests that models may inadvertently learn to recognize baseline states or moods rather than the intended emotional responses. With this in mind,

the emotion classification model should take into consideration incorrect annotations, either by detecting and ignoring them, or by employing strategies to mitigate their impact.

8.2.3 Emotion Classification

Physiological signals have intrinsic characteristics that introduce challenges in emotion recognition. The signals are inherently noisy and little movements can introduce artefacts. Moreover, the signals are highly variable across individuals for physiological reasons. Thirdly, there is a diverse set of sensor modalities, and even for an individual modality, the same sensor can be read in different body locations or be developed with different technical characteristics, modifying its data. One further challenge in emotion recognition is the inherent variability in how emotions are experienced and expressed by different individuals, where the same emotion can be expressed differently either by cultural, or environmental factors, resulting in subject-dependent physiological responses and meanings for self-reported annotations.

The heterogeneity in physiological data complicates the creation of uniform datasets, hindering their transfer and comparability across varied experimental setups. This issue is exacerbated by the scarcity of data in emotion recognition, primarily due to the high costs and time requirements for gathering large amounts of data in laboratory settings, where most physiological data are collected. Overall, these constraints limit the training of advanced models and the development of robust and generalizable emotion recognition algorithms, such as transformers, which require large, comprehensive datasets to be trained effectively.

Lastly, a persistent issue in machine learning is the imbalance of class labels. Such imbalance hinders effective model learning, especially with limited data, skewing results towards more frequently represented classes.

8.2.4 Moving Forward

Reflecting on these insights, it is evident that the field of affective computing stands at a crossroads, with significant opportunities for advancement as well as substantial challenges to overcome. Future research must prioritize how to deal with individual variability in self-reporting and physiological responses, and the development of robust, multi-modal data collection methods that encompass the full spectrum of emotional experience. Equally, innovative annotation methodologies that accurately capture the subjective and fleeting nature of emotions, correct them and distinguish them from longer-lasting moods, are essential.

Addressing data scarcity and imbalance requires a concerted effort to create larger, more diverse datasets. Such datasets should aim for standardisation across sensor modalities and data collection

protocols to facilitate comparative studies and the application of advanced machine learning models. Moreover, exploring synthetic data generation and augmentation techniques could offer pathways to mitigate data limitations.

The journey through this thesis highlights the complexity of emotions, and its multidisciplinary nature, requiring developments in physiological data acquisition, annotation and analysis to capture the intricacies of emotion recognition through computational means. As the field progresses, more generalisable and naturalistic datasets, a holistic, nuanced understanding of emotions, paired with advanced computational and annotation techniques, will be paramount in unlocking the full potential of affective computing.

8.2.5 Affective Computing Applications

Affective computing has the promise of revolutionising human-computer interaction, significantly improving lives across healthcare, entertainment, education, and beyond.

1. **Arts & Entertainment:** Affective computing can facilitate interactive narratives by capturing the audience's emotional responses to adjust content in real-time, without explicit user inputs. Similarly, live performances can create unique experiences through real-time emotional feedback between performers and audiences, enhancing the collective engagement and impact of the art.
2. **Neuro-marketing:** Affective computing can refine marketing strategies by uncovering subconscious consumer preferences and enabling more targeted advertising and content creation. Emotional analysis offers valuable insights for tailoring advertisements, music, and games to individual responses, paving the way for personalized content recommendations across platforms.
3. **Healthcare:** With mental health issues affecting a significant portion of the global population, affective computing wearable devices can aim for early detection, diagnosis, and monitoring of conditions such as anxiety and depression related to chronic negative emotions. Personal affective monitors could identify stress triggers or beneficial practices, supporting mental well-being in daily life, high-stress professions, or therapeutic settings.
4. **e-Learning:** The shift towards online learning platforms, accelerated by the COVID-19 pandemic, underscores the potential of the digital medium in education. By adjusting teaching methods to match students' attention, motivation, and comprehension levels, affective computing can enhance the effectiveness of digital learning experiences.

Overall, the applications of affective computing are vast, promising to improve millions of lives globally from enhancing the depth of artistic experiences to supporting mental health and personalising entertainment and education.

8.3 Future Work

Emotion recognition is a field that has been growing in the last few years, with the development of new algorithms and the collection of new datasets. With this in mind, in Chapter 3 the literature algorithms and features were evaluated on publicly available datasets to understand the current status of the field. The experimental results evidenced that the performance of the algorithms is highly dependent on the dataset and its characteristics. Upon these results, future work could focus on transfer learning between datasets, the development of a methodology that could be applied across the different dataset setups, or improving the explainability of the models, to understand the features that are being used to classify the emotions.

Next, in Chapter 4, the addition of group context was explored in comparison to intrasubject data for emotion recognition classification. In this context, future work may tackle limitations of the proposed method, namely the requirement of annotated labels at test time, or the exploration of alternative annotations instead of the external annotations by experts, which may not be related to the true underlying emotional experiences. Lastly, this work relies on either dyads or groups of 4, although groups can vary widely in size. This limitation arises from the available datasets but should be considered.

To meet this limitation, in Chapter 5, the EmotiphAI platform was developed to collect physiological data from groups. Future work could focus on the development of a more robust and comfortable form factor for the wearable device, towards a device with no cables and evaluate the usability of the wearable and acquisition platform.

To annotate the physiological data, Chapter 6 introduced and validated the EmotiphAI Annotator, a web-based tool that allows the annotation of emotions in long-duration content, retrospectively. Future work may tackle the limitations of the proposed method, namely the study of the relation between the EDA events and their emotional meaningfulness, and the exploration of diverse scene moments for annotation.

In the final chapter, Chapter 7, the G-REx, a novel dataset that features physiological data collected from groups in naturalistic settings was created. Future directions could include augmenting this dataset with environmental and behavioural metrics such as humidity, room temperature, and activity data; investigating alternative annotation methods, such as image-based annotations for capturing facial expressions or actions like talking and eating; and applying advanced deep learning techniques that account for the temporal dynamics and multimodal nature of emotional data, like LSTMs or one-dimensional-CNN, or the weight of different moments of the data like Transformers, towards a complete view of emotions and the improvement of its classification.

Appendices

Contents

Appendix A	Benchmarking Emotion Recognition	171
Appendix B	Group Emotion Recognition	176
Appendix C	Data Acquisition	183

Appendix A Benchmarking Emotion Recognition

Table A .1: Features label and description across sensor modality: PPG, ECG, EDA, and respiration; and domains: Non-linear, Spectral, Temporal and Statistical.

Domain	Name	Description
PPG	ons	Signal onsets.
	hr	Heart rate.
ECG	rpeaks	R-peak location indices.
	nni	NN intervals in ms or s.
	hr	Instantaneous Heart rate in bpm.
	pnn20	Percentage of successive NN intervals that differ by more than 20 ms.
	pnn50	Percentage of successive NN intervals that differ by more than 50 ms.
	sdu_index	Mean of the standard deviations of all the NN intervals for each 5 min segment.
	sdu	Standard deviation of the average NN intervals for each 5 min segment.
	fft_peak_VLF	Peak frequencies of the very low frequency bands [0.00Hz - 0.04Hz] in Hz.
	fft_peak_LF	Peak frequencies of the low frequency bands [0.04Hz - 0.15Hz] in Hz.
	fft_peak_HF	Peak frequencies of the high frequency bands [0.15Hz - 0.40Hz] in Hz.
	fft_abs_VLF	Absolute power of the very low-frequency bands [0.00Hz - 0.04Hz].
	fft_abs_LF	Absolute power of the low-frequency bands [0.04Hz - 0.15Hz].
	fft_abs_HF	Absolute power of the high-frequency bands [0.15Hz - 0.40Hz].
	fft_rel_VLF	Relative power of the very low-frequency bands [0.00Hz - 0.04Hz].
	fft_rel_LF	Relative power of the low-frequency bands [0.04Hz - 0.15Hz].
	fft_rel_HF	Relative power of the high-frequency bands [0.15Hz - 0.40Hz].
fft_log_VLF	Log power of the very low-frequency bands [0.00Hz - 0.04Hz].	

Continued on next page

Table A .1 continued from previous page

Domain	Name	Description
	fft_log_LF	Log power of the low-frequency bands [0.04Hz - 0.15Hz].
	fft_log_HF	Log power of the high-frequency bands [0.15Hz - 0.40Hz].
	fft_total	Total power over all frequency bands.
	fft_ratio	Ratio of LF to HF.
EDA	onsets	Signal EDR events onsets.
	pks	Signal EDR events peaks.
	amps	Signal EDR events amplitudes.
	phasic_rate	Signal EDR events rate in 60s.
	rise_ts	Rise times, i.e. onset-peak time difference.
	half_rise	Half Rise times, i.e. the time between onset and 50% amplitude.
	half_rec	Half Recovery times, i.e. the time between peak and 63% amplitude.
	six_rise	63% rise times, i.e. time between onset and 63% amplitude.
	six_rec	63% recovery times, i.e. time between peak and 63% amplitude.
	onPkVol	EDR onset-peaks volume.
	pkOnVol	EDR peaks-onsets volume.
	EDRVolRatio	Ratio between On-Pk and Pk-On volumes.
Respiration	zeros	Signal zero crossing indexes.
	hr	Respiration rate.
	inhale	Inhalation volume.
	exhale	Exhalation volume.
	inhExhRatio	Ratio between Inhalation and Exhalation.
	inhale_dur	Inhalation time duration.
	exhale_dur	Exhalation time duration.
Non-Linear	sd1	Standard deviation of the major axis in the Poincaré Plot.
	sd2	Standard deviation of the minor axis in the Poincaré Plot.
	sd12	Ratio between SD2 and SD1 (SD2/SD1).
	poincare	Area of the Poincaré Plot fitted ellipse.
	sample_entropy	Sample entropy of the NNI series.
	dfa_alpha1	Alpha value of the short-term Detrended Fluctuation Analysis.
	dfa_alpha2	Alpha value of the long-term Detrended Fluctuation Analysis.
	tinn_n	N value of the TINN computation.
	tinn_m	M value of the TINN computation.
	tinn	Baseline width of the NNI histogram based on the triangular Interpolation.
	triangular_index	Ratio of the total number of NN intervals to the height of the histogram.
Spectral	spectral_maxpeaks	Number of peaks in the spectrum signal.
	spect_var	Amount of the variation of the spectrum across time.
	curve_distance	Euclidean distance between the spectrum cumulative sum and evenly spaced signal length.
	spectral_roll_off	Frequency so 95% of the signal energy is below that value.
	spectral_roll_on	Frequency so 5% of the signal energy is below that value.

Continued on next page

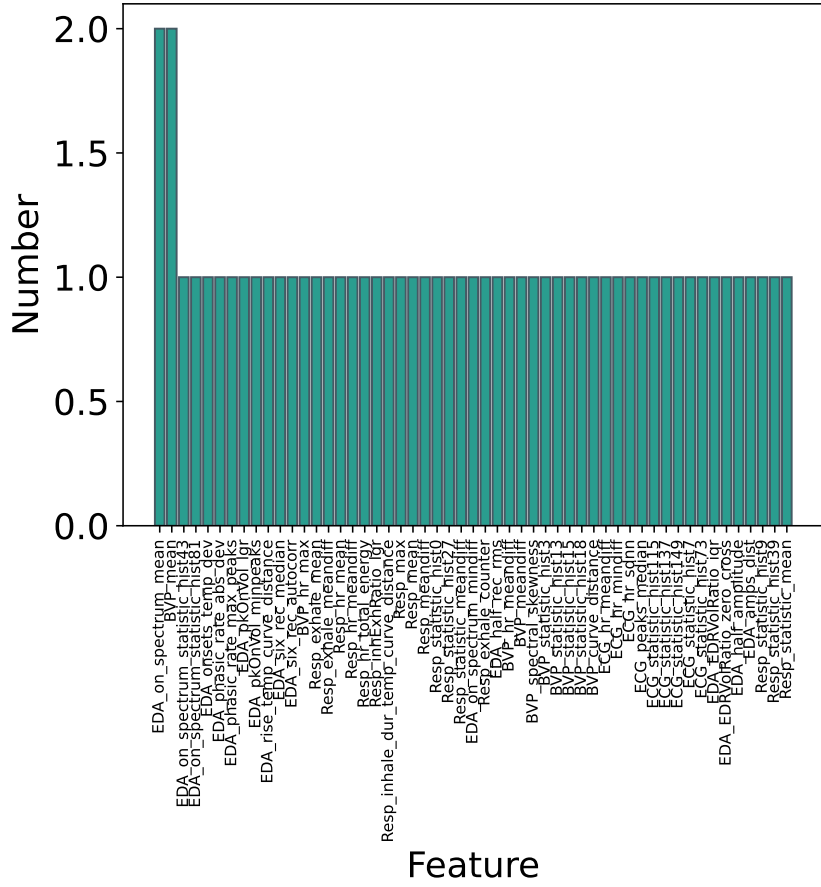
Table A .1 continued from previous page

Domain	Name	Description
	spectral_dec	Amount of decreasing in the spectral amplitude.
	spectral_slope	Amount of decreasing in the spectral amplitude.
	spectral_centroid	Centroid of the signal spectrum.
	spectral_spread	Variance of the signal spectrum i.e. how it spreads around its mean value.
	spectral_kurtosis	Kurtosis of the signal spectrum i.e. describes the flatness of the spectrum distribution.
	spectral_skewness	Skewness of the signal spectrum i.e. describes the asymmetry of the spectrum distribution.
	max_frequency	Maximum frequency of the signal spectrum maximum amplitude.
	fundamental_frequency	Fundamental frequency of the signal.
	max_power_spectrum	Spectrum maximum value.
	mean_power_spectrum	Spectrum mean value.
	spectral_hist	Histogram of the signal spectrum.
Statistic	mean	Mean of the signal.
	median	Median of the signal.
	var	Signal variance.
	std	Signal standard deviation.
	abs_dev	Absolute signal deviation.
	kurtosis	Signal kurtosis.
	skewness	Signal skewness.
	iqr	Interquartile Range.
	meanadev	Mean absolute deviation.
	medadev	Median absolute deviation.
	rms	Root Mean Square.
	statistic_hist	Histogram.
Temporal	maxAmp	Signal maximum amplitude.
	minAmp	Signal minimum amplitude.
	max	Signal max value.
	min	Signal min value.
	dist	Length of the signal.
	autocorr	Signal autocorrelation.
	zero_cross	Number of times the sinal crosses the zero axis.
	meanadiff	Mean absolute differences.
	mindiff	Minimum differences.
	maxdiff	Maximum differences.
	sadiff	Sum of absolute differences.
	meandiff	Mean of differences.
	meddiff	Median of differences.
	temp_centroid	Temporal centroid.
	total_energy	Total energy.

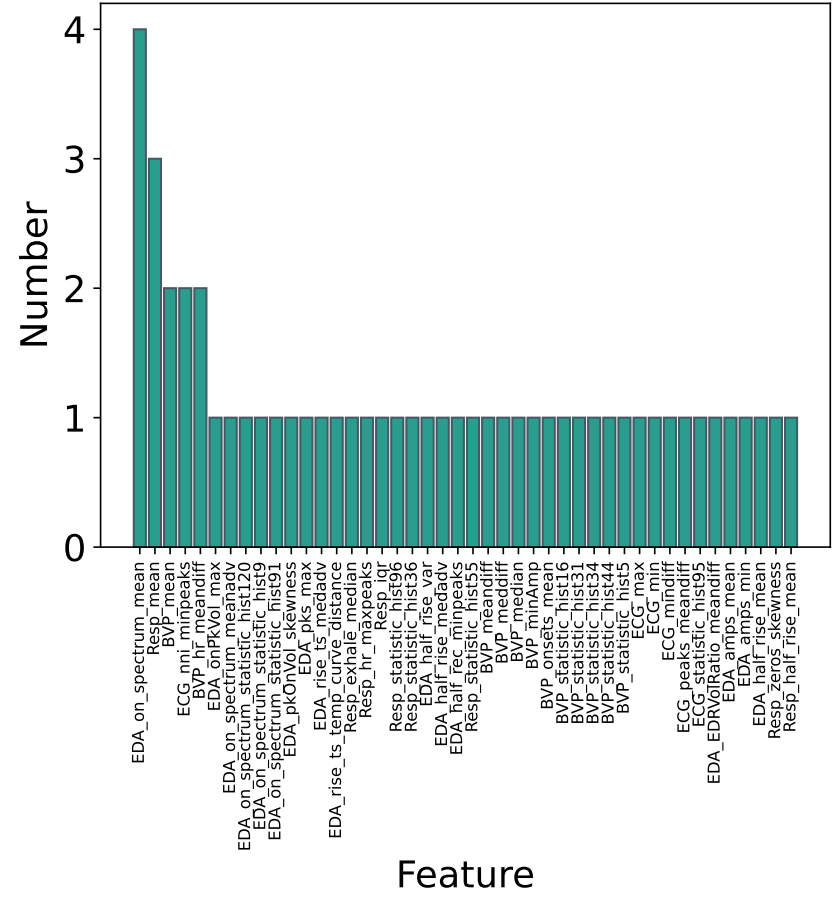
Continued on next page

Table A .1 continued from previous page

Domain	Name	Description
	minpeaks	Number of minimum peaks.
	maxpeaks	Number of maximum peaks.
	temp_dev	Temporal deviation.
	counter	Length of the signal.
	temp_curve_distance	Euclidean distance of the signal's cumulative sum to the respective linear regression.
	temp_curve_distance_	Volume of the signal's cumulative sum to the respective linear regression.
	vol	



(a) Arousal



(b) Valence

Figure 1: Summary of the features selected in the feature selection step, combining all the datasets.

Appendix B Group Emotion Recognition

Table B 1: Features extracted from EDA and RR-interval signals.

Domain	Name	Description
EDA	len_pks	Peak count
	pks_amp	Mean peak amplitude
	rise_ts	Mean rise time
	sum_pks_amp	Sum of peak amplitudes
	sum_rise_ts	Sum of rise times
	sum_areas	Sum of areas under the curve
	sum_area5s	Area under curve within a time response window of 5 sec after each stimulus onset
	mean_EDA	Mean Electrodermal Activity
	std_EDA	Standard deviation of EDA
	kurtosis_EDA	Kurtosis of EDA
	skew_EDA	Skewness of EDA
	mean_1sder	Mean first derivative of EDA
	mean_neg_1sder	Mean negative first derivative of EDA
	GSR_respEnergy	Galvanic Skin Response response energy
	sum_spec	Sum of spectral power in the [0-0.5] Hz band
	spectPower_var	Variance of spectral power
	EDR_area	Signal magnitude area of Electrodermal Response
	spect_kurt	Spectrum kurtosis
	mobility	Mobility of EDA
	complexity	Complexity of EDA
	zeroCross	Number of zero crossings
	mfcc_kurt	Mel-Frequency Cepstral Coefficients kurtosis
	mfcc_skew	MFCC skewness
	mfcc_mean	Mean MFCC
	mfcc_std	Standard deviation of MFCC
	mfcc_median	Median MFCC
HRV	nni_counter	Number of normal-to-normal intervals (NNIs)
	hr_mean	Mean heart rate
	hr_min	Minimum heart rate
	hr_max	Maximum heart rate
	hr_std	Standard deviation of heart rate
	nni_diff_mean	Mean of the differences between adjacent NNIs
	nni_diff_min	Minimum of the differences between adjacent NNIs
	nni_diff_max	Maximum of the differences between adjacent NNIs
	sdnn	Standard deviation of NNIs
	sdnn_index	Mean of the standard deviations of all NNIs for all 5-minute segments of the entire recording
	sdann	Standard deviation of the average NNIs for all 5-minute segments of the entire recording
	rmssd	Root mean square of successive differences between NNIs

Continued on next page

Table B 1 continued from previous page

Domain	Name	Description
	sdsd	Standard deviation of successive differences between NNIs
	nn50	Number of pairs of successive NNIs that differ by more than 50 ms
	pnn50	Proportion of nn50 divided by total number of NNIs
	nn20	Number of pairs of successive NNIs that differ by more than 20 ms
	pnn20	Proportion of nn20 divided by total number of NNIs
	tinn_n	Baseline width of the triangular interpolation of the highest peak of the NN interval histogram (minimum value)
	tinn_m	Baseline width of the triangular interpolation of the highest peak of the NN interval histogram (maximum value)
	tinn	Baseline width of the triangular interpolation of the highest peak of the NN interval histogram
	tri_index	Triangular index
	fft_peak_VLF	Peak of the very low frequency band in FFT
	fft_peak_LF	Peak of the low frequency band in FFT
	fft_peak_HF	Peak of the high frequency band in FFT
	fft_abs_VLF	Absolute power of the very low frequency band in FFT spectral analysis
	fft_abs_LF	Absolute power of the low frequency band in FFT spectral analysis
	fft_abs_HF	Absolute power of the high frequency band in FFT spectral analysis
	fft_rel_VLF	Relative power of the very low frequency band in FFT spectral analysis
	fft_rel_LF	Relative power of the low frequency band in FFT spectral analysis
	fft_rel_HF	Relative power of the high frequency band in FFT spectral analysis
	fft_log_VLF	Logarithmic power of the very low frequency band in FFT spectral analysis
	fft_log_LF	Logarithmic power of the low frequency band in FFT spectral analysis
	fft_log_HF	Logarithmic power of the high frequency band in FFT spectral analysis
	fft_norm_LF	Normalized power of the low frequency band in FFT spectral analysis
	fft_norm_HF	Normalized power of the high frequency band in FFT spectral analysis
	fft_ratio	Ratio of the low to high frequency power in FFT spectral analysis
	fft_total	Total power across all frequency bands in FFT spectral analysis
	lomb_peak_VLF	Peak power of the very low frequency band in Lomb-Scargle spectral analysis
	lomb_peak_LF	Peak power of the low frequency band in Lomb-Scargle spectral analysis
	lomb_peak_HF	Peak power of the high frequency band in Lomb-Scargle spectral analysis
	lomb_abs_VLF	Absolute power of the very low frequency band in Lomb-Scargle spectral analysis
	lomb_abs_LF	Absolute power of the low frequency band in Lomb-Scargle spectral analysis
	lomb_abs_HF	Absolute power of the high frequency band in Lomb-Scargle spectral analysis
	lomb_rel_VLF	Relative power of the very low frequency band in Lomb-Scargle spectral analysis
	lomb_rel_LF	Relative power of the low frequency band in Lomb-Scargle spectral analysis
	lomb_rel_HF	Relative power of the high frequency band in Lomb-Scargle spectral analysis
	lomb_log_VLF	Logarithmic power of the very low frequency band in Lomb-Scargle spectral analysis
	lomb_log_LF	Logarithmic power of the low frequency band in Lomb-Scargle spectral analysis
	lomb_log_HF	Logarithmic power of the high frequency band in Lomb-Scargle spectral analysis
	lomb_norm_LF	Normalized power of the low frequency band in Lomb-Scargle spectral analysis
	lomb_norm_HF	Normalized power of the high frequency band in Lomb-Scargle spectral analysis

Continued on next page

Table B 1 continued from previous page

Domain	Name	Description
	lomb_ratio	Ratio of the low to high frequency power in Lomb-Scargle spectral analysis
	lomb_total	Total power of the Lomb-Scargle periodogram
	ar_peak_VLF	Peak of the very low frequency band in AR spectrum
	ar_peak_LF	Peak of the low frequency band in AR spectrum
	ar_peak_HF	Peak of the high frequency band in AR spectrum
	ar_abs_VLF	Absolute power of the very low frequency band in autoregressive (AR) spectral analysis
	ar_abs_LF	Absolute power of the low frequency band in autoregressive (AR) spectral analysis
	ar_abs_HF	Absolute power of the high frequency band in autoregressive (AR) spectral analysis
	ar_rel_VLF	Relative power of the very low frequency band in autoregressive (AR) spectral analysis
	ar_rel_LF	Relative power of the low frequency band in autoregressive (AR) spectral analysis
	ar_rel_HF	Relative power of the high frequency band in autoregressive (AR) spectral analysis
	ar_log_VLF	Logarithmic power of the very low frequency band in autoregressive (AR) spectral analysis
	ar_log_LF	Logarithmic power of the low frequency band in autoregressive (AR) spectral analysis
	ar_log_HF	Logarithmic power of the high frequency band in autoregressive (AR) spectral analysis
	ar_norm_LF	Normalized power of the low frequency band in autoregressive (AR) spectral analysis
	ar_norm_HF	Normalized power of the high frequency band in autoregressive (AR) spectral analysis
	ar_ratio	Ratio of low frequency band power to high frequency band power in autoregressive (AR) spectral analysis
	ar_total	Total power of all frequency bands in autoregressive (AR) spectral analysis
	sd1	Poincaré plot standard deviation perpendicular to the line of identity
	sd2	Poincaré plot standard deviation along the line of identity
	sd_ratio	Ratio of sd1 to sd2
	ellipse_area	Area of the ellipse fitted to the Poincaré plot
	sampen	Sample entropy
	dfa_alpha1	Short-term scaling exponent of detrended fluctuation analysis
	dfa_alpha2	Long-term scaling exponent of detrended fluctuation analysis

Table B 2: Results for the WGS on the AMIGOS dataset for the Arousal dimension. The results are shown in terms of accuracy (Acc), M-F1, computation time per sample (Time), and weights standard deviation (weight STD). The best result is shown in bold. ED – Euclidean Distance

Data	Similarity Metric	Acc (%)	W-F1-score (%)	M-F1-score (%)	Time (s)	Weight STD
EDR	Pearson	54.28 ± 04.55	58.81 ± 05.67	48.81 ± 06.05	0.008 ± 0.001	0.15 ± 0.00
	Spearman	54.34 ± 04.20	58.88 ± 05.46	48.77 ± 05.93	0.011 ± 0.002	0.15 ± 0.00
	Cosine	59.02 ± 03.75	63.09 ± 03.73	52.34 ± 06.77	0.009 ± 0.001	0.13 ± 0.01
	DTW	82.14 ± 05.46	82.00 ± 06.77	71.11 ± 10.81	0.011 ± 0.002	0.16 ± 0.02
	Euclidean Distance	82.39 ± 05.48	82.21 ± 06.78	71.39 ± 10.80	0.009 ± 0.001	0.13 ± 0.02
	Cross-Correlation (Max)	82.02 ± 05.48	81.73 ± 06.92	70.62 ± 10.64	0.009 ± 0.001	0.20 ± 0.02
	Recurrence Plot (ED)	82.52 ± 05.36	82.37 ± 06.47	71.61 ± 10.16	0.217 ± 0.020	0.15 ± 0.01
	Coherence (Sum)	82.17 ± 05.02	82.02 ± 06.24	71.16 ± 09.80	0.010 ± 0.001	0.19 ± 0.01
EDL	Pearson	57.49 ± 03.13	61.75 ± 03.77	51.24 ± 06.22	0.009 ± 0.001	0.11 ± 0.01
	Spearman	57.33 ± 03.22	61.59 ± 03.84	51.09 ± 06.19	0.012 ± 0.003	0.11 ± 0.01
	Cosine	55.18 ± 07.51	59.47 ± 08.21	49.52 ± 08.09	0.011 ± 0.005	0.09 ± 0.02
	DTW	82.49 ± 05.52	82.41 ± 06.57	71.73 ± 10.45	0.014 ± 0.004	0.18 ± 0.01
	Euclidean Distance	82.67 ± 05.56	82.58 ± 06.61	72.02 ± 10.43	0.009 ± 0.002	0.16 ± 0.02
	Cross-Correlation (Max)	79.87 ± 06.60	80.18 ± 07.26	69.28 ± 09.42	0.012 ± 0.005	0.28 ± 0.03
	Recurrence Plot (ED)	82.56 ± 05.33	82.29 ± 06.65	71.57 ± 09.80	0.209 ± 0.006	0.16 ± 0.00
	Coherence (Sum)	82.02 ± 05.31	81.83 ± 06.73	70.98 ± 10.13	0.012 ± 0.001	0.20 ± 0.00
EDA – FV	Pearson	82.57 ± 05.67	82.20 ± 07.15	71.33 ± 10.80	0.003 ± 0.003	0.05 ± 0.01
	Spearman	82.43 ± 05.69	82.09 ± 07.17	71.15 ± 10.81	0.003 ± 0.000	0.07 ± 0.01
	Cosine	82.64 ± 05.62	82.22 ± 07.18	71.29 ± 10.87	0.002 ± 0.000	0.02 ± 0.00
	Euclidean Distance	82.60 ± 05.58	82.23 ± 07.07	71.29 ± 10.88	0.004 ± 0.001	0.06 ± 0.01
HRV – FV	Pearson	82.60 ± 05.58	82.23 ± 07.07	71.29 ± 10.88	0.002 ± 0.000	0.06 ± 0.01
	Spearman	82.65 ± 05.29	82.36 ± 06.78	71.47 ± 10.68	0.003 ± 0.000	0.09 ± 0.02
	Cosine	82.75 ± 05.45	82.38 ± 06.94	71.51 ± 10.75	0.003 ± 0.001	0.00 ± 0.00
	Euclidean Distance	82.74 ± 05.30	82.38 ± 06.79	71.46 ± 10.67	0.004 ± 0.002	0.05 ± 0.01
LR HRV – FV	Pearson	76.86 ± 10.43	77.72 ± 10.36	67.07 ± 10.61	0.010 ± 0.002	0.11 ± 0.04
	Spearman	78.32 ± 10.44	78.85 ± 10.31	68.28 ± 10.85	0.010 ± 0.002	0.08 ± 0.07
	Cosine	83.01 ± 05.18	82.76 ± 06.39	71.99 ± 10.42	0.009 ± 0.002	0.09 ± 0.03
	Euclidean Distance	83.07 ± 04.92	82.87 ± 06.05	72.15 ± 10.11	0.007 ± 0.002	0.12 ± 0.05
	Average Pooling	82.65 ± 05.49	82.26 ± 07.05	71.34 ± 10.78	0.010 ± 0.005	0.00 ± 0.00

Table B 3: Results for the WGS for the AMIGOS dataset, Valence dimension.

Data	Similarity Metric	Acc (%)	W-F1-score (%)	M-F1-score (%)	Time (s)	Weights STD
EDR	Pearson	53.83 ± 03.52	55.35 ± 03.52	52.26 ± 03.96	0.009 ± 0.002	0.15 ± 0.00
	Spearman	54.08 ± 03.59	55.60 ± 03.60	52.47 ± 04.09	0.010 ± 0.002	0.15 ± 0.00
	Cosine	57.21 ± 04.02	58.64 ± 03.57	55.32 ± 04.84	0.009 ± 0.002	0.13 ± 0.01
	DTW	82.24 ± 06.35	82.69 ± 05.90	80.50 ± 07.45	0.010 ± 0.001	0.16 ± 0.02
	Euclidean Distance	82.61 ± 06.25	83.06 ± 05.79	80.90 ± 07.44	0.010 ± 0.002	0.13 ± 0.02
	Cross-Correlation (Max)	82.25 ± 06.10	82.67 ± 05.68	80.45 ± 07.17	0.011 ± 0.000	0.20 ± 0.02
	Recurrence Plot (ED)	82.30 ± 06.00	82.75 ± 05.58	80.58 ± 07.09	0.219 ± 0.046	0.15 ± 0.01
	Coherence (Sum)	82.29 ± 05.99	82.72 ± 05.55	80.53 ± 07.08	0.012 ± 0.001	0.19 ± 0.01
EDL	Pearson	57.51 ± 03.15	58.90 ± 03.08	55.67 ± 03.78	0.009 ± 0.001	0.11 ± 0.01
	Spearman	57.44 ± 03.05	58.83 ± 03.03	55.56 ± 03.65	0.011 ± 0.001	0.11 ± 0.01
	Cosine	54.62 ± 07.15	56.04 ± 06.94	52.96 ± 07.24	0.008 ± 0.001	0.09 ± 0.02
	DTW	82.04 ± 06.25	82.48 ± 05.83	80.28 ± 07.31	0.010 ± 0.001	0.18 ± 0.01
	Euclidean Distance	82.24 ± 06.39	82.68 ± 05.97	80.51 ± 07.45	0.008 ± 0.000	0.16 ± 0.02
	Cross-Correlation (Max)	79.16 ± 05.82	79.59 ± 05.62	77.14 ± 06.43	0.009 ± 0.001	0.28 ± 0.03
	Recurrence Plot (ED)	82.61 ± 06.14	83.05 ± 05.71	80.88 ± 07.31	0.226 ± 0.044	0.16 ± 0.00
	Coherence (Sum)	81.97 ± 05.95	82.38 ± 05.60	80.15 ± 07.00	0.011 ± 0.003	0.20 ± 0.00
EDA – FV	Pearson	82.48 ± 06.46	82.96 ± 06.03	80.81 ± 07.65	0.004 ± 0.003	0.05 ± 0.01
	Spearman	82.46 ± 06.40	82.93 ± 05.96	80.77 ± 07.60	0.004 ± 0.001	0.07 ± 0.01
	Cosine	82.72 ± 06.55	83.19 ± 06.11	81.10 ± 07.68	0.003 ± 0.001	0.02 ± 0.00
	Euclidean Distance	82.80 ± 06.52	83.26 ± 06.09	81.16 ± 07.63	0.004 ± 0.006	0.06 ± 0.01
HRV – FV	Pearson	82.24 ± 06.34	82.73 ± 05.90	80.61 ± 07.42	0.003 ± 0.001	0.09 ± 0.01
	Spearman	82.29 ± 06.39	82.78 ± 05.96	80.67 ± 07.48	0.005 ± 0.001	0.09 ± 0.02
	Cosine	82.64 ± 06.53	83.12 ± 06.08	81.03 ± 07.65	0.003 ± 0.001	0.02 ± 0.00
	Euclidean Distance	82.67 ± 06.48	83.15 ± 06.03	81.07 ± 07.57	0.004 ± 0.002	0.05 ± 0.01
LR HRV – FV	Pearson	77.06 ± 09.93	77.67 ± 09.55	75.01 ± 11.01	0.005 ± 0.001	0.10 ± 0.04
	Spearman	77.26 ± 09.93	77.88 ± 09.55	75.23 ± 11.09	0.007 ± 0.001	0.05 ± 0.07
	Cosine	82.75 ± 06.23	83.22 ± 05.78	81.11 ± 07.37	0.005 ± 0.001	0.08 ± 0.03
	Euclidean Distance	82.48 ± 06.36	82.95 ± 05.92	80.82 ± 07.44	0.007 ± 0.000	0.11 ± 0.04
	Average Pooling	82.70 ± 06.46	83.19 ± 06.01	81.11 ± 07.58	0.008 ± 0.001	0.00 ± 0.00

Table B 4: Results for the WGS for the KEemoCon dataset, Arousal dimension.

Data	Similarity Metric	Acc (%)	W-F1-score (%)	M-F1-score (%)	Time (s)
EDR	Pearson	54.19 ± 09.31	60.14 ± 11.68	43.73 ± 8.97	0.000 ± 0.000
	Spearman	54.62 ± 10.36	60.39 ± 12.67	43.85 ± 09.70	0.001 ± 0.000
	Cosine	62.51 ± 13.63	66.41 ± 16.84	46.64 ± 09.72	0.000 ± 0.000
	Cross-Correlation (Max)	70.39 ± 21.21	70.63 ± 21.91	47.34 ± 12.61	0.001 ± 0.000
EDL	Pearson	58.46 ± 09.07	63.81 ± 11.72	46.41 ± 10.18	0.001 ± 0.000
	Spearman	58.38 ± 08.91	63.72 ± 11.80	46.19 ± 09.72	0.002 ± 0.000
	Cosine	67.40 ± 12.20	69.64 ± 14.46	49.23 ± 10.85	0.000 ± 0.000
	Cross-Correlation (Max)	70.24 ± 12.32	71.58 ± 15.05	50.37 ± 09.57	0.001 ± 0.000
EDA – FV	Pearson	70.87 ± 22.50	70.80 ± 22.90	52.63 ± 21.28	0.000 ± 0.000
	Spearman	70.74 ± 22.52	70.69 ± 22.96	52.52 ± 21.31	0.000 ± 0.000
	Cosine	70.87 ± 22.50	70.80 ± 22.90	52.63 ± 21.28	0.000 ± 0.000
HRV – FV	Pearson	70.37 ± 21.89	70.62 ± 22.52	47.55 ± 13.40	0.000 ± 0.000
	Spearman	70.55 ± 22.44	70.58 ± 22.92	52.43 ± 21.28	0.000 ± 0.000
	Cosine	70.87 ± 22.50	70.80 ± 22.90	52.63 ± 21.28	0.000 ± 0.000
LR HRV – FV	Pearson	61.45 ± 17.33	65.17 ± 19.34	47.86 ± 12.17	0.008 ± 0.012
	Spearman	61.33 ± 17.16	65.01 ± 19.21	47.51 ± 11.95	0.011 ± 0.015
	Cosine	67.08 ± 25.04	67.36 ± 25.39	48.88 ± 20.47	0.009 ± 0.012
	Average Pooling	70.87 ± 22.50	70.80 ± 22.90	52.63 ± 21.28	0.000 ± 0.000

Table B 5: Results for the WGS for the K-EmoCon dataset, Valence dimension.

Data	Similarity Metric	Acc (%)	W-F1-score (%)	M-F1-score (%)	Time (s)
EDR	Pearson	87.04 ± 12.20	88.12 ± 12.45	61.20 ± 21.34	0.001 ± 0.000
	Spearman	87.00 ± 12.05	88.19 ± 12.24	61.42 ± 21.34	0.001 ± 0.000
	Cosine	86.83 ± 10.66	88.05 ± 11.72	58.73 ± 19.36	0.000 ± 0.000
	Cross-Correlation (Max)	89.82 ± 09.74	89.91 ± 10.92	59.53 ± 19.38	0.001 ± 0.000
EDL	Pearson	87.51 ± 10.20	88.29 ± 11.23	57.68 ± 19.62	0.000 ± 0.001
	Spearman	87.60 ± 10.12	88.35 ± 11.19	57.79 ± 19.64	0.001 ± 0.000
	Cosine	89.39 ± 09.35	89.72 ± 11.13	62.03 ± 21.03	0.000 ± 0.000
	Cross-Correlation (Max)	90.04 ± 08.91	90.16 ± 10.63	65.09 ± 22.55	0.001 ± 0.000
EDA – FV	Pearson	90.04 ± 09.88	90.04 ± 10.94	64.90 ± 22.59	0.000 ± 0.000
	Spearman	90.04 ± 09.75	90.09 ± 10.84	65.01 ± 22.65	0.000 ± 0.000
	Cosine	90.04 ± 09.88	90.04 ± 10.94	64.90 ± 22.59	0.0001 ± 0.000
HRV – FV	Pearson	90.05 ± 09.65	90.05 ± 10.89	64.91 ± 22.55	0.000 ± 0.000
	Spearman	90.00 ± 09.79	90.03 ± 10.89	64.92 ± 22.59	0.001 ± 0.000
	Cosine	90.04 ± 09.88	90.04 ± 10.94	64.90 ± 22.59	0.000 ± 0.000
LR HRV – FV	Pearson	88.35 ± 10.36	88.93 ± 11.68	62.41 ± 22.07	0.002 ± 0.002
	Spearman	88.35 ± 10.21	88.96 ± 11.56	62.54 ± 22.12	0.004 ± 0.002
	Cosine	88.92 ± 10.32	89.19 ± 11.55	61.74 ± 22.38	0.004 ± 0.010
	Average Pooling	90.04 ± 09.88	90.04 ± 10.94	64.90 ± 22.59	0.000 ± 0.000

Table B 6: Hyperparameters space values used in the WGS and intrapersonal approach. Nomenclature: Dimension space (Dim) in arousal (A), and valence (V); Grace period (Grace Per.); Learning rate(Learn. Rate); Weight decay (Weight Dec.). The parameters are shown per dataset for AMIGOS (AM) and K-EmoCon (K).

	Dim.	EDA	Feature HRV	Spect.	Image RP	Morphology EDA-EDR-EDL
Batch size	A V	16 (K, AM), 128 (K, AM)	16 (K, AM), 128 (K, AM)	16 (K), 128 (K, AM)	16 (K), 128 (K, AM)	16 (K), 128 (K), 256 (AM) 16 (K), 256 (K, A)
Epoch	A V	800 (K, AM) 600 (K), 800 (AM)	800 (K, AM) 600 (K), 800 (AM)	300 (K), 800 (AM) 60 (K), 800 (AM)	800 (AM), 1000 (K) 60 (K), 800 (AM)	800 (K), 1000 (AM)
Gamma	A V	0.5 (AM), 1.5 (K)	0.5 (AM), 1.5 (K)	0.5 (K) 1.5 (K)	1.5 (K)	1.5 (K)
Grace Per.	A V	0 (AM), 10 (K), 50 (K) 0 (AM), 10 (K), 100 (K)	10 (K), 50 (K), 60 (AM) 0 (A), 10 (K), 100 (K)	0 (K), 10 (AM), 20 (K) 0 (K), 10 (AM), 50 (K)	0 (K, AM), 50 (K) 0 (K), 10 (A), 40 (K)	10 (K), 100 (K), 900 (AM) 10 (K), 200 (K), 900 (AM)
Learn. Rate	A V	1e-3 (K, AM), 1e-5 (K, A)	1e-3 (K, AM), 1e-5 (K, AM)	1e-3 (K), 1e-5 (K), 1e-6 (AM) 1e-6 (K, AM), 1e-3 (K)	1e-5 (K, AM), 1e-3 (K) 1e-6 (K, A), 1e-3 (K)	1e-5 (K, AM), 1e-3 (K)
Patience	A V	6 (AM), 10 (K)	6 (AM), 10 (K)	5 (AM), 10 (K)	5 (AM), 10 (K)	10 (K), 50 (AM)
Weight Dec.	A V			0.01 (K, AM) 0.01 (K, AM)		

Appendix C Data Acquisition

Table C 1: Technical specifications for the Zigbee, Bluetooth and WiFi protocols¹²³.

	Zigbee	Bluetooth (BLE)	WiFi
Network type	LAN	PAN	WPAN, WLAN, WWAN, WMAN
Topology	Self-Forming, Self-Healing MESH	Mesh and Star	Mesh, Star, Hybrid, P2P, Bus, Ring
Network Protocol	Zigbee PRO 2015		802.11n
Radio Technology	IEEE 802.15.4-2011	IEEE 802.15.1	IEEE 802.11
Frequency Band / Channels	2.4 GHz (ISM band) 16-channels (2 MHz wide)	2.402 GHz to 2.48 GHz	2.4 GHz and the 5 GHz bands
Data Rate	250 Kbits/sec	270 Kbits/sec	450Mbps
Communication Range (Average)	75-100m indoor	77m indoors	100m indoors
Theoretical # of nodes	Up to 65,000	32 767	
Modulation	DSSS	FHSS	OFDM, MIMO
Transmit power	10 mW	10 mW	100mW

Table C 2: Technical specifications of the TP-Link Wireless N 450Mbps (TL-WR940N)⁴ and TP-Link MR3020 3G/Wi-Fi⁵.

	TL-MR3020	TL-WR940N
Standards	IEEE 802.11n, IEEE 802.11g, IEEE 802.11b	IEEE 802.11n/b/g 2.4 GHz
Protocols	IPv4, IPv6	IPv4, IPv6
Frequencies	2.4 GHz	2.4 GHz
Antenna	1 internal antenna	3 antennas
Data Rate	300Mbps	450 Mbps (802.11n)
Transmit Power	< 20dBm	< 20dBm

¹<https://intel.com/content/www/us/en/support/articles/000005725/wireless/legacy-intel-wireless-products.html>; Accessed: 20/02/2024

²https://standards.ieee.org/standard/802_15_4-2011.html; Accessed: 20/02/2024

³<https://ieee802.org/15/pub/TG1.html>; Accessed: 20/02/2024

⁴[static.tp-link.com/2018/201810/20181022/TL-WR940N\(EU\)6.0-datasheet.pdf](https://static.tp-link.com/2018/201810/20181022/TL-WR940N(EU)6.0-datasheet.pdf); Accessed on 20/02/2024

⁵[static.tp-link.com/2017/201712/20171207/TL-MR3020\(EU\)_3.20.pdf](https://static.tp-link.com/2017/201712/20171207/TL-MR3020(EU)_3.20.pdf); Accessed on 20/02/2024

References

- [1] G. Coppin and D. Sander, "Theoretical approaches to emotion and its measurement," in *Emotion Measurement* (H. L. Meiselman, ed.), pp. 3–30, Woodhead Publishing, 2016.
- [2] R. W. Picard, "Affective computing-MIT media laboratory perceptual computing section technical report no. 321," *Cambridge, MA*, vol. 2139, p. 92, 1995.
- [3] A. Ortony and G. L. Clore, "Can an appraisal model be compatible with psychological constructionism," in *The Psychological Construction of Emotion*, pp. 305–333, New York: Guilford Press, 2015.
- [4] D. Sander, "Models of emotion," *The Cambridge handbook of human affective neuroscience*, pp. 5–56, 2013.
- [5] I. B. Mauss and M. D. Robinson, "Measures of emotion: A review," *Cognition and Emotion*, vol. 23, no. 2, pp. 209–237, 2009.
- [6] F. Larradet, R. Niewiadomski, G. Barresi, D. G. Caldwell, and L. S. Mattos, "Toward emotion recognition from physiological signals in the wild: Approaching the methodological issues in real-life data collection," *Frontiers in Psychology*, vol. 11, p. 1111, 2020.
- [7] P. Bota, T. Zhang, A. El Ali, A. Fred, H. P. da Silva, and P. Cesar, "Group synchrony for emotion recognition using physiological signals," *IEEE Trans. on Affective Computing*, vol. 14, no. 4, pp. 2614–2625, 2023.
- [8] P. Bota, R. Silva, C. Carreiras, A. Fred, and H. P. da Silva, "Biosppy: A python toolbox for physiological signal processing," *SoftwareX*, vol. 26, p. 101712, 2024.
- [9] P. Bota, C. Wang, A. L. N. Fred, and H. Plácido da Silva, "A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals," *IEEE Access*, vol. 7, no. 1, pp. 140990–141020, 2019.
- [10] P. Bota, C. Wang, A. Fred, and H. Silva, "Emotion assessment using feature fusion and decision fusion classification based on physiological data: Are we there yet?," *Sensors*, vol. 20, no. 17, p. 4723, 2020.
- [11] P. Bota, A. Fred, J. Valente, C. Wang, and H. P. da Silva, "A dissimilarity-based approach to automatic classification of biosignal modalities," *Applied Soft Computing*, vol. 115, p. 108203, 2022.
- [12] P. Bota, E. Flety, H. P. d. Silva, and A. Fred, "EmotiphAI: a biocybernetic engine for real-time biosignals acquisition in a collective setting," *Neural Computing and Applications*, vol. 35, no. 8, pp. 5721–5736, 2023.
- [13] R. Silva, G. Salvador, P. Bota, A. Fred, and H. Plácido da Silva, "Impact of sampling rate and interpolation on photoplethysmography and electrodermal activity signals' waveform morphology and feature extraction," *Neural Computing and Applications*, vol. 35, no. 8, pp. 5661–5677, 2023.
- [14] M. N. Supelnic, A. F. Ferreira, P. Bota, L. Brás-Rosário, and H. Plácido da Silva, "Benchmarking of sensor configurations and measurement sites for out-of-the-lab photoplethysmography," *Sensors*, vol. 24, no. 1, 2024.
- [15] P. Bota, P. Cesar, A. Fred, and H. Silva, "Exploring retrospective annotation in long-videos for emotion recognition," *IEEE Trans. on Affective Computing*, vol. 15, no. 3, pp. 1–12, 2024.
- [16] P. Bota, J. Brito, A. Fred, P. Cesar, and H. Silva, "A real-world dataset of group emotion experiences based on physiological data," *Scientific Data*, vol. 11, no. 1, pp. 1–17, 2024.

- [17] L. Aly, L. Godinho, P. Bota, G. Bernardes, and H. P. da Silva, "Acting emotions: A comprehensive dataset of elicited emotions," *Scientific Data*, 2024.
- [18] P. Bota, C. Wang, A. L. Fred, and H. Silva, "A wearable system for electrodermal activity data acquisition in collective experience assessment," in *Proc. of the Int'l Conf. on Enterprise Information Systems*, pp. 606–613, 2020.
- [19] G. F. D. Salvador, P. Bota, V. Vinayagamoorthy, H. Plácido da Silva, and A. Fred, "Smartphone-based content annotation for ground truth collection in affective computing," in *Proc. of the Int'l Conf. on Interactive Media Experiences*, p. 199–204, ACM, 2021.
- [20] P. Bota, A. Fred, J. Valente, and H. Silva, "Automatic classification of physiological signals modalities," in *Int'l Meeting of the Portuguese Society of Physiology*, 2019.
- [21] L. Aly, P. Bota, L. Godinho, G. Bernardes, and H. Silva, "Acting emotions: physiological correlates of emotional valence and arousal dynamics in theatre," in *ACM Int'l Conf. on Interactive Media Experiences*, pp. 381–386, 2022.
- [22] S. Morgado, "AI-powered emotion well-being for everyone," MSc thesis, Instituto Superior Técnico da Universidade de Lisboa, 2024.
Advisor: Sangra Gama (IST-UL/INESC-ID) and Hugo P. da Silva (IST-UL/IT).
- [23] J. Alves, "Facial emotion recognition for mental well-being assessment in the workplace," MSc thesis, Instituto Superior Técnico da Universidade de Lisboa, 2022.
Advisor: Hugo P. da Silva (IST-UL/IT) and Ana Fred (IST-UL/IT).
- [24] G. Salvador, "Real world group emotional analytics using electrodermal activity signals," Master's thesis, Instituto Superior Técnico, Universidade de Lisboa, 2021.
- [25] C. Lima, "Psychophysiological effects of guided imagery based intervention on the academic development of children," MSc thesis, Faculdade de Ciências da Universidade de Lisboa, 2024.
Advisor: Hugo P. da Silva (IST-UL/IT), Nuno Matela (FCUL) and Brigida Ferreira (FCUL).
- [26] T. Talento, "Movie emotional content analysis," 2023. BSc final project, Instituto Superior Técnico - Universidade de Lisboa
Co-Advisor: Hugo Silva (IST-UL/IT).
- [27] I. Salema, "Emotion analysis using facial expression recognition," 2023. BSc final project, Instituto Superior Técnico - Universidade de Lisboa
Co-Advisor: Hugo Silva (IST-UL/IT).
- [28] M. Supelnic, "Benchmarking of sensor configurations and measurement sites for out-of-the-lab photoplethysmography," 2024. BSc final project, Instituto Superior Técnico - Universidade de Lisboa
Co-Advisor: Afonso Ferreira (INESC), Hugo Silva (IST-UL/IT).
- [29] F. Silva and Z. Xu, 2024. BSc final project, Instituto Superior Técnico - Universidade de Lisboa
Co-Advisor: Hugo Silva (IST-UL/IT).
- [30] S. Silvestre, "Learning enhancement using affective computing," 2024. BSc final project, Instituto Superior Técnico - Universidade de Lisboa
Co-Advisor: Hugo Silva (IST-UL/IT), Cátia Costa (HLuz), José Moreira (HLuz).
- [31] J. Brito, "A proof-of-concept neurosecurity study based on peripheral physiological signals," 2022 – 2023. Technical Report, Instituto Telecomunicações.
- [32] A. Gonçalves, "Integration of ScientISST core in EmotiphAI," 2022. Instituto Telecomunicações.
- [33] P. Correia, "Real-world data collection," 2021 – 2022. Instituto Telecomunicações.
- [34] C. Bento, "Data collection using the FMCI device," 2019 - 2020. Instituto Telecomunicações.

- [35] D. Venancio, "Kubernetes for group data collection - EmotiphAI," 2023. Instituto Telecomunicações.
- [36] F. A. Santos, "Firmware and hardware developer - ScientISST," 2023 - 2024. Instituto Telecomunicações.
- [37] P. Bota and d. S. H. P., "Emotion assessment in the wild," in *Modern Technologies Enabling Innovative Methods for Maritime Monitoring and Strengthening Resilience in Maritime Critical Infrastructures*, (Lisbon, Portugal), Jan 2024.
- [38] P. Bota and A. Ferreira, "What do your biosignals say about you," in *INBIO (Introdução aos Bio-sinais)*, (Lisbon, Portugal), Jul 2023.
- [39] P. Bota, M. Abreu, and H. P. da Silva, "Introduction to machine learning and applications," in *Exercise Prescription and Health Promotion*, (Instituto Politécnico de Leiria, Leiria, Portugal), Feb 2020.
- [40] P. Bota and J. Brito, "EmotiphAI: How to detect emotions in a group of people," in *Maker Faire*, (Lisbon, Portugal), Feb 2023.
- [41] P. Bota, H. P. da Silva, and M. Abreu, "Introduction to biosignal acquisition," in *1st Int'l Meeting of the Portuguese Society of Physiology*, (Lisbon, Portugal), Oct 2019.
- [42] P. Bota and M. Abreu, "Signal processing and machine learning," in *summer course of CEiiA (Centre of Engineering and Product Development)*, (Instituto Superior Técnico, Lisbon, Portugal), Jul 2020.
- [43] P. Bota and M. Abreu, "Introduction to machine learning and applications," in *Clynx*, (online), Jul 2020.
- [44] P. Bota and M. Abreu, "Physiological signal classification," in *Congress of the Brazilian Society of Physiology*, (online), Sep 2020.
- [45] P. Bota, A. S. Carmo, M. Abreu, and S. Monteiro, "EmotiphAI and IT group projects showcase." Noite Europeia dos Investigadores, Sep 2023.
- [46] P. Bota and A. M., "EmotiphAI and IT group projects showcase." Dia Internacional das Raparigas nas Tecnologias de Informação e Comunicação, Apr 2023.
- [47] P. Bota, J. Brito, R. Silva, and V. Garção, "Biosignals acquisition and visualisation." FIC.A (International Festival of Science), Oct 2022.
- [48] P. Bota, S. Monteiro, L. Pereira, R. Silva, and R. Maciel, "EmotiphAI and it group projects showcase." 22nd anniversary of Campus Taguspark, Nov 2022.
- [49] P. Bota and M. Abreu, "Dissemination of FMCI xhuanet signal acquisition devices." Web Summit, Lisbon, Portugal, Nov 2019.
In collaboration with Xhuanet FMCI.
- [50] A. M. Schmitter, "17th and 18th Century Theories of Emotions," in *The Stanford Encyclopedia of Philosophy* (E. N. Zalta, ed.), Metaphysics Research Lab, Stanford University, Summer 2021 ed., 2021.
- [51] B. Fehr and J. A. Russell, "Concept of emotion viewed from a prototype perspective," *Journal of Experimental Psychology: General*, vol. 113, no. 3, p. 464, 1984.
- [52] C. Darwin, *The Expression of the Emotions in Man and Animals*. Cambridge Library Collection - Darwin, Evolution and Genetics, Cambridge University Press, 2013.
- [53] P. N. Juslin and P. Laukka, "Communication of emotions in vocal expression and music performance: Different channels, same code?," *Psychological Bulletin*, vol. 129, no. 5, p. 770, 2003.
- [54] A. Celeghin, M. Diano, A. Bagnis, M. Viola, and M. Tamietto, "Basic emotions in human neuroscience: Neuroimaging and beyond," *Frontiers in Psychology*, vol. 8, 2017.
- [55] L. Barrett, *How Emotions Are Made: The Secret Life of the Brain*. HarperCollins, 2017.
- [56] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *Journal of Personality and Social Psychology*, vol. 17 2, pp. 124–9, 1971.

- [57] R. Plutchik, "A psychoevolutionary theory of emotions," *Social Science Information*, vol. 21, no. 4-5, pp. 529–553, 1982.
- [58] C. E. Izard, "Basic emotions, natural kinds, emotion schemas, and a new paradigm," *Perspectives on Psychological Science*, vol. 2, no. 3, pp. 260–280, 2007.
- [59] S. H. Shmurak, "Demystifying emotion: Introducing the affect theory of Silvan Tomkins to objectivists," *The Journal of Ayn Rand Studies*, vol. 8, no. 1, pp. 1–18, 2006.
- [60] M. Kowalska and M. Wróbel, "Basic emotions," in *Encyclopedia of personality and individual differences*, pp. 377–382, Springer, 2020.
- [61] K. Oatley and P. N. Johnson-laird, "Towards a cognitive theory of emotions," *Cognition and Emotion*, vol. 1, no. 1, pp. 29–50, 1987.
- [62] P. Ekman and D. Cordaro, "What is meant by calling emotions basic," *Emotion Review*, vol. 3, no. 4, pp. 364–370, 2011.
- [63] R. W. Levenson, "Basic emotion questions," *Emotion Review*, vol. 3, no. 4, pp. 379–386, 2011.
- [64] J. Panksepp and D. Watt, "What is basic about basic emotions? lasting lessons from affective neuroscience," *Emotion Review*, vol. 3, no. 4, pp. 387–396, 2011.
- [65] J. Cottingham, R. Stoothoff, and D. Murdoch, *The Passions of the Soul*, vol. 1, p. 325–404. Cambridge University Press, 1985.
- [66] M. B. Arnold, "Emotion and personality. vol. i. psychological aspects," 1960.
- [67] P. Ekman, W. V. Friesen, and J. Hager, "Environmental psychology & nonverbal behavior," *Environmental Psychology and Nonverbal Behavior*, vol. 1, no. 1, pp. 56–75, 1976.
- [68] J. Posner, J. A. Russel, and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Development and Psychopathology*, vol. 17, no. 3, p. 715–734, 2005.
- [69] R. Lazarus, "Progress on a cognitive-motivational-relational theory of emotion," *The American Psychologist*, vol. 46, no. 8, p. 819–834, 1991.
- [70] A. Moors, "Theories of emotion causation: A review," *Cognition and Emotion*, vol. 23, no. 4, pp. 625–662, 2009.
- [71] L. F. Barrett, "2 - Navigating the science of emotion," in *Emotion Measurement* (H. L. Meiselman, ed.), pp. 31–63, Woodhead Publishing, 2016.
- [72] K. R. Scherer, E. Clark-Polner, and M. Mortillaro, "In the eye of the beholder? universality and cultural specificity in the expression and perception of emotion," *Int'l Journal of Psychology*, vol. 46, no. 6, pp. 401–435, 2011.
- [73] J. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, pp. 1161–1178, 1980.
- [74] E. Clark-Polner, T. D. Wager, A. B. Satpute, and L. F. Barrett, "Neural fingerprinting: Meta-analysis, variation and the search for brain-based essences in the science of emotion," *The Handbook of Emotion*, pp. 146–65, 2016.
- [75] J. A. Russell, "Core affect and the psychological construction of emotion," *Psychological Review*, vol. 110, no. 1, pp. 145–172, 2003.
- [76] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "Emotion, motivation, and anxiety: brain mechanisms and psychophysiology," *Biological Psychiatry*, vol. 44, no. 12, pp. 1248–1263, 1998.
- [77] R. Thayer, *The origin of everyday moods: Managing energy, tension and stress*. Oxford University Press, 1989.
- [78] D. Watson, D. Wiese, J. G. Vaidya, and A. Tellegen, "The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence," *Journal of Personality and Social Psychology*, vol. 76, pp. 820–838, 1999.

- [79] L. F. Barrett and J. A. Russell, "The structure of current affect: Controversies and emerging consensus," *Current Directions in Psychological Science*, vol. 8, no. 1, pp. 10–14, 1999.
- [80] L. F. Barrett, "Discrete emotions or dimensions? the role of valence focus and arousal focus," *Cognition and Emotion*, vol. 12, no. 4, pp. 579–599, 1998.
- [81] P. Lang, M. Greenwald, M. Bradley, and A. Hamm, "Looking at pictures: Affective, facial, visceral, and behavioral reactions," *Psychophysiology*, vol. 30, no. 3, pp. 261–273, 1993.
- [82] P. R. Kleinginna and A. M. Kleinginna, "A categorized list of emotion definitions, with suggestions for a consensual definition," *Motivation and Emotion*, vol. 5, no. 4, pp. 345–379, 1981.
- [83] K. R. Scherer, "What are emotions? and how can they be measured?," *Social science information*, vol. 44, no. 4, pp. 695–729, 2005.
- [84] R. Jacob-Dazarola, J. C. Ortíz Nicolás, and L. Cárdenas Bayona, "5 - behavioral measures of emotion," in *Emotion Measurement* (H. L. Meiselman, ed.), pp. 101–124, Woodhead Publishing, 2016.
- [85] R. W. Levenson, "The autonomic nervous system and emotion," *Emotion Review*, vol. 6, no. 2, pp. 100–112, 2014.
- [86] D. L. Paulhus, S. Vazire, *et al.*, "The self-report method," *Handbook of Research Methods in Personality Psychology*, vol. 1, no. 2007, pp. 224–239, 2007.
- [87] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras, "AMIGOS: A dataset for affect, personality and mood research on individuals and groups," *IEEE Trans. on Affective Computing*, vol. 12, no. 2, pp. 479–493, 2018.
- [88] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a multimodal dataset for wearable stress and affect detection," in *Proc. of the ACM Int'l Conf. on multimodal interaction*, pp. 400–408, 2018.
- [89] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis using physiological signals," *IEEE Trans. on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2011.
- [90] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.
- [91] D. Watson and L. A. Clark, "The PANAS-X: Manual for the positive and negative affect schedule-expanded form," 1994.
- [92] M. M. Rahman, A. K. Sarkar, M. A. Hossain, M. S. Hossain, M. R. Islam, M. B. Hossain, J. M. Quinn, and M. A. Moni, "Recognition of human emotions using EEG signals: A review," *Computers in Biology and Medicine*, vol. 136, p. 104696, 2021.
- [93] M. Li and B.-L. Lu, "Emotion classification based on gamma-band EEG," in *Annual Int'l Conf. of the IEEE Engineering in Medicine and Biology Society*, pp. 1223–1226, 2009.
- [94] R. J. Davidson, P. Ekman, C. D. Saron, J. A. Senulis, and W. V. Friesen, "Approach-withdrawal and cerebral asymmetry: emotional expression and brain physiology: I," *Journal of Personality and Social Psychology*, vol. 58, no. 2, p. 330, 1990.
- [95] S. K. Sutton and R. J. Davidson, "Prefrontal brain asymmetry: A biological substrate of the behavioral approach and inhibition systems," *Psychological Science*, vol. 8, no. 3, pp. 204–210, 1997.
- [96] F. C. Murphy, I. Nimmo-Smith, and A. D. Lawrence, "Functional neuroanatomy of emotions: a meta-analysis," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 3, pp. 207–233, 2003.
- [97] I. Daly, D. Williams, F. Hwang, A. Kirke, E. R. Miranda, and S. J. Nasuto, "Electroencephalography reflects the activity of sub-cortical brain regions during approach-withdrawal behaviour while listening to music," *Scientific Reports*, vol. 9, no. 1, p. 9415, 2019.
- [98] H. Saarimäki, A. Gotsopoulos, I. P. Jääskeläinen, J. Lampinen, P. Vuilleumier, R. Hari, M. Sams, and L. Nummenmaa, "Discrete Neural Signatures of Basic Emotions," *Cerebral Cortex*, vol. 26, no. 6, pp. 2563–2573, 2015.

- [99] M. Lévêque and M. Lévêque, "The neuroanatomy of emotions," *Psychosurgery: New Techniques for Brain Disorders*, pp. 49–106, 2014.
- [100] R. Singh, R. Sharma, V. Chauhan, and K. Chatterjee, "Neurobiological underpinnings of emotions," *Industrial Psychiatry Journal*, vol. 30, no. Suppl 1, p. S308, 2021.
- [101] I. Daum, H. J. Markowitsch, and M. Vandekerckhove, *Neurobiological Basis of Emotions*, pp. 111–138. Springer US, 2009.
- [102] S. Gu, F. Wang, C. Cao, E. Wu, Y.-Y. Tang, and J. H. Huang, "An integrative way for studying neural basis of basic emotions with fMRI," *Frontiers in Neuroscience*, vol. 13, p. 628, 2019.
- [103] S. Hamann, "Mapping discrete and dimensional emotions onto the brain: controversies and consensus," *Trends in Cognitive Sciences*, vol. 16, no. 9, pp. 458–466, 2012.
- [104] K. A. Lindquist, T. D. Wager, H. Kober, E. Bliss-Moreau, and L. F. Barrett, "The brain basis of emotion: a meta-analytic review," *Behavioral and Brain Sciences*, vol. 35, no. 3, pp. 121–143, 2012.
- [105] S. Tomkins, *Affect imagery consciousness: Volume I: The positive affects*. Springer publishing company, 1962.
- [106] Y.-I. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 97–115, 2001.
- [107] K. R. Scherer, "Vocal communication of emotion: A review of research paradigms," *Speech Communication*, vol. 40, no. 1-2, pp. 227–256, 2003.
- [108] B. App, D. McIntosh, C. Reed, and M. Hertenstein, "Nonverbal channel use in communication of emotion: How may depend on why," *Emotion (Washington, D.C.)*, vol. 11, pp. 603–17, 2011.
- [109] I. B. Mauss, R. W. Levenson, L. McCarter, F. H. Wilhelm, and J. J. Gross, "The tie that binds? coherence among emotion experience, behavior, and physiology," *Emotion*, vol. 5, no. 2, p. 175, 2005.
- [110] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biological Psychology*, vol. 84, no. 3, pp. 394–421, 2010.
- [111] B. H. Friedman, "Feelings and the body: The Jamesian perspective on autonomic specificity of emotion," *Biological Psychology*, vol. 84, no. 3, pp. 383–393, 2010.
- [112] G. Stemmler, "Physiological processes during emotion," *The Regulation of Emotion*, vol. 415, pp. 33–70, 2004.
- [113] K. A. Lindquist, E. H. Siegel, K. S. Quigley, and L. F. Barrett, "The hundred-year emotion war: are emotions natural kinds or psychological constructions? comment on lench, flores, and bench (2011)," *Psychological Bulletin*, vol. 139, no. 1, pp. 255–263, 2013.
- [114] J. Cacioppo, G. Berntson, J. Larsen, K. Poehlmann, and T. Ito, *The Psychophysiology of Emotion*, pp. 173–191. Guilford Press, 2000.
- [115] E. H. Siegel, M. K. Sands, W. V. den Noortgate, P. Condon, Y. Chang, J. G. Dy, K. S. Quigley, and L. F. Barrett, "Emotion fingerprints or emotion populations? a meta-analytic investigation of autonomic features of emotion categories," *Psychological Bulletin*, vol. 144, p. 343–393, 2018.
- [116] K. Quigley and L. Barrett, "Is there consistency and specificity of autonomic changes during emotional episodes? guidance from the conceptual act theory and psychophysiology," *Biological Psychology*, vol. 98, pp. 82–94, 2014.
- [117] S. for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures, W. Boucsein, D. C. Fowles, S. Grimnes, G. Ben-Shakhar, W. T. Roth, M. E. Dawson, and D. L. Fillion, "Publication recommendations for electrodermal measurements," *Psychophysiology*, vol. 49, no. 8, pp. 1017–1034, 2012.
- [118] E. Babaei, B. Tag, T. Dingler, and E. Velloso, "A critique of electrodermal activity practices at chi," in *Proc. of the CHI Conf. on Human Factors in Computing Systems*, pp. 1–14, 2021.

- [119] W. Boucsein, *Principles of Electrodermal Phenomena*, pp. 1–86. Springer US, 2012.
- [120] G. J. Norman, E. Necka, and G. G. Berntson, “4 - The psychophysiology of emotions,” in *Emotion Measurement* (H. L. Meiselman, ed.), pp. 83–98, Woodhead Publishing, 2016.
- [121] D. Castaneda, A. Esparza, M. Ghamari, C. Soltanpur, and H. Nazeran, “A review on wearable photoplethysmography sensors and their potential future applications in health care,” *Int'l Journal of Biosensors & Bioelectronics*, vol. 4, no. 4, p. 195, 2018.
- [122] T. Tamura and Y. Maeda, *Photoplethysmogram*, pp. 159–192. Springer Int'l Publishing, 2018.
- [123] E. D. Chan, M. M. Chan, and M. M. Chan, “Pulse oximetry: Understanding its basic principles facilitates appreciation of its limitations,” *Respiratory Medicine*, vol. 107 6, pp. 789–99, 2013.
- [124] J.-M. Fernández-Dols and M.-A. Ruiz-Belda, “Are smiles a sign of happiness? gold medal winners at the olympic games,” *Journal of Personality and Social Psychology*, vol. 69, no. 6, p. 1113, 1995.
- [125] G. A. Van Kleef and A. H. Fischer, “Emotional collectives: How groups shape emotions and emotions shape groups,” *Cognition and Emotion*, vol. 30, no. 1, pp. 3–19, 2016.
- [126] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, “A multimodal database for affect recognition and implicit tagging,” *IEEE Trans. on Affective Computing*, vol. 3, no. 1, pp. 42–55, 2011.
- [127] R. Subramanian, J. Wache, M. K. Abadi, R. L. Vieriu, S. Winkler, and N. Sebe, “ASCERTAIN: Emotion and personality recognition using commercial sensors,” *IEEE Trans. on Affective Computing*, vol. 9, no. 2, pp. 147–160, 2018.
- [128] J. A. Healey, *Wearable and automotive systems for affect recognition from physiology*. PhD thesis, Massachusetts Institute of Technology, 2000.
- [129] S. Carvalho, J. Leite, S. Galdo-Álvarez, and Ó. F. Gonçalves, “The emotional movie database (EMDB): A self-report and psychophysiological study,” *Applied Psychophysiology and Biofeedback*, vol. 37, no. 4, pp. 279–294, 2012.
- [130] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, and N. Sebe, “DECAF: MEG-based multimodal database for decoding affective physiological responses,” *IEEE Trans. on Affective Computing*, vol. 6, no. 3, pp. 209–222, 2015.
- [131] K. Sharma, C. Castellini, F. Stulp, and E. L. van den Broek, “Continuous, real-time emotion annotation: A novel joystick-based analysis framework,” *IEEE Trans. on Affective Computing*, vol. 11, no. 1, pp. 78–84, 2020.
- [132] V. Markova, T. Ganchev, and K. Kalinkov, “CLAS: A database for cognitive load, affect and stress recognition,” in *Proc. Int'l Conf. on Biomedical Innovations and Applications*, pp. 1–4, 2019.
- [133] Y. Liu, T. Gedeon, S. Caldwell, S. Lin, and Z. Jin, “Emotion recognition through observer’s physiological signals,” *arXiv preprint arXiv:2002.08034*, 2020.
- [134] T. Xue, A. E. Ali, T. Zhang, G. Ding, and P. Cesar, “CEAP-360VR: A continuous physiological and behavioral emotion annotation dataset for 360° videos,” *IEEE Trans. on Multimedia*, pp. 1–1. 10.1109/TMM.2021.3124080 (2021).
- [135] J. Pinto, “Exploring physiological multimodality for emotional assessment,” Master’s thesis, Instituto Superior Técnico, 2019.
- [136] K. Kutt, D. Drażyk, L. Żuchowska, M. Szelążek, S. Bobek, and G. Nalepa, “BIRAFFE2, a multimodal dataset for emotion-based personalization in rich affective game environments,” vol. 9, no. 1, p. 274, 2022. Scientific Data <https://doi.org/10.1038/s41597-022-01402-6> (2022).
- [137] C. Park, N. Cha, S. Kang, A. Kim, A. Khandoker, L. Hadjileontiadis, A. Oh, Y. Jeong, and U. Lee, “K-EmoCon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations,” *Scientific Data*, vol. 7, no. 1, p. 293, 2020.
- [138] A. Zlatintsi, P. Koutras, G. Evangelopoulos, N. Malandrakis, N. Efthymiou, K. Pastra, A. Potamianos, and P. Maragos, “COGNIMUSE: A multimodal video database annotated with saliency, events, semantics and emotion with application to summarization,” *EURASIP Journal on Image and Video Processing*, no. 1, p. 54, 2017.

- [139] Y. Baveye, E. Dellandréa, C. Chamaret, and L. Chen, "Deep learning vs. kernel methods: Performance for emotion prediction in videos," in *Proc. of the Int'l Conf. on Affective Computing and Intelligent Interaction*, pp. 77–83, 2015.
- [140] X. Shui, M. Zhang, Z. Li, X. Hu, F. Wang, and D. Zhang, "A dataset of daily ambulatory psychological and physiological recording for emotion research." Synapse <https://doi.org/10.7303/syn22418021> (2021).
- [141] R. Cowie, E. Douglas-Cowie, S. Savvidou, E. McMahon, M. Sawey, and M. Schröder, "FEELTRACE: An instrument for recording perceived emotion in real time," in *ISCA Tutorial and Research Workshop on Speech and Emotion*, 2000.
- [142] R. Cowie, M. Sawey, C. Doherty, J. Jaimovich, C. Fyans, and P. Stapleton, "GTrace: General trace program compatible with EmotionML," in *Humaine Association Conf. on Affective Computing and Intelligent Interaction*, pp. 709–710, 2013.
- [143] G. A. van Kleef, *Emotion: An interpersonal perspective*, p. 1–10. Studies in Emotion and Social Interaction, Cambridge University Press, 2016.
- [144] J. I. Menges and M. Kilduff, "Group emotions: Cutting the gordian knots concerning terms, levels of analysis, and processes," *Academy of Management Annals*, vol. 9, no. 1, pp. 845–928, 2015.
- [145] W. McDougall, *Revival: An Outline of Psychology (1968)*. Routledge, 2018.
- [146] E. Durkheim, "The elementary forms of religious life," in *Social Theory Re-wired*, pp. 52–67, Routledge, 2016.
- [147] D. Páez, B. Rimé, N. Basabe, A. Włodarczyk, and L. Zumeta, "Psychosocial effects of perceived emotional synchrony in collective gatherings," *Journal of Personality and Social Psychology*, vol. 108, no. 5, p. 711, 2015.
- [148] J. R. Kelly and S. G. Barsade, "Mood and emotions in small groups and work teams," *Organizational behavior and human decision processes*, vol. 86, no. 1, pp. 99–130, 2001.
- [149] A. Goldenberg, D. Garcia, E. Halperin, and J. J. Gross, "Collective emotions," *Current Directions in Psychological Science*, vol. 29, no. 2, pp. 154–160, 2020.
- [150] S. G. Barsade and D. E. Gibson, "Group emotion: A view from top and bottom," 1998.
- [151] S. G. Barsade and A. P. Knight, "Group affect," *Annu. Rev. Organ. Psychol. Organ. Behav.*, vol. 2, no. 1, pp. 21–46, 2015.
- [152] E. Hatfield, J. T. Cacioppo, and R. L. Rapson, "Primitive emotional contagion," 1992.
- [153] R. S. Lazarus, *Emotion and adaptation*. Oxford University Press, 1991.
- [154] E. Hatfield, R. L. Rapson, and Y.-C. L. Le, "Emotional contagion and empathy," *The Social Neuroscience of Empathy*, p. 19, 2011.
- [155] J. M. George, "Personality, affect, and behavior in groups," *Journal of Applied Psychology*, vol. 75, no. 2, p. 107, 1990.
- [156] P. B. Barger and A. A. Grandey, "Service with a smile and encounter satisfaction: Emotional contagion and appraisal mechanisms," *Academy of Management Journal*, vol. 49, no. 6, pp. 1229–1238, 2006.
- [157] T. Sy and J. N. Choi, "Contagious leaders and followers: Exploring multi-stage mood contagion in a leader activation and member propagation model," *Organizational Behavior and Human Decision Processes*, vol. 122, no. 2, pp. 127–140, 2013.
- [158] A. D. Kramer, J. E. Guillory, and J. T. Hancock, "Experimental evidence of massive-scale emotional contagion through social networks," *Proc. of the National Academy of Sciences of the United States of America*, vol. 111, no. 24, p. 8788, 2014.
- [159] R. Ilies, D. T. Wagner, and F. P. Morgeson, "Explaining affective linkages in teams: Individual differences in susceptibility to contagion and individualism-collectivism," *Journal of Applied Psychology*, vol. 92, no. 4, p. 1140, 2007.
- [160] P. Schmidt, A. Reiss, R. Duerichen, and K. Van Laerhoven, "Wearable affect and stress recognition: A review," *arXiv preprint arXiv:1811.08854*, 2018.
- [161] L. Shu, J. Xie, M. Yang, Z. Li, Z. Li, D. Liao, X. Xu, and X. Yang, "A review of emotion recognition using physiological signals," *Sensors*, vol. 18, no. 7, p. 2074, 2018.

- [162] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis," *IEEE Trans. on Affective Computing*, vol. 1, no. 2, pp. 81–97, 2010.
- [163] A. Samara, M. L. R. Menezes, and L. Galway, "Feature extraction for emotion recognition and modelling using neurophysiological data," in *Proc. of the Int'l Conf. on ubiquitous computing and communications and Int'l symposium on cyberspace and security*, pp. 138–144, IEEE, 2016.
- [164] J. Zhang, M. Chen, S. Hu, Y. Cao, and R. Kozma, "PNN for EEG-based emotion recognition," in *IEEE Proc. of the Int'l Conf. on Systems, Man, and Cybernetics*, pp. 002319–002323, IEEE, 2016.
- [165] P. Gong, H. T. Ma, and Y. Wang, "Emotion recognition based on the multiple physiological signals," in *IEEE Prof. of the Int'l Conf. on Real-time Computing and Robotics*, pp. 140–143, IEEE, 2016.
- [166] G. K. Verma and U. S. Tiwary, "Multimodal fusion framework: A multiresolution approach for emotion classification and recognition from physiological signals," *NeuroImage*, vol. 102, pp. 162–172, 2014.
- [167] V. Kolodyazhnyi, S. D. Kreibig, J. J. Gross, W. T. Roth, and F. H. Wilhelm, "An affective computing approach to physiological emotion specificity: Toward subject-independent and stimulus-independent classification of film-induced emotions," *Psychophysiology*, vol. 48, no. 7, pp. 908–922, 2011.
- [168] D. Shin, D. Shin, and D. Shin, "Development of emotion recognition interface using complex EEG/ECG bio-signal for interactive contents," *Multimedia Tools and Applications*, vol. 76, pp. 11449–11470, 2017.
- [169] F. Agrafioti, D. Hatzinakos, and A. K. Anderson, "ECG pattern analysis for emotion detection," *IEEE Trans. on Affective Computing*, vol. 3, no. 1, pp. 102–115, 2011.
- [170] W. Wen, G. Liu, N. Cheng, J. Wei, P. Shangguan, and W. Huang, "Emotion recognition based on multi-variant correlation of physiological signals," *IEEE Trans. on Affective Computing*, vol. 5, no. 2, pp. 126–140, 2014.
- [171] J. Kim and E. André, "Emotion recognition based on physiological changes in music listening," *IEEE Trans. on pattern analysis and machine intelligence*, vol. 30, no. 12, pp. 2067–2083, 2008.
- [172] C. Zong and M. Chetouani, "Hilbert-huang transform based physiological signals analysis for emotion recognition," in *IEEE Int'l symposium on signal processing and information technology*, pp. 334–339, IEEE, 2009.
- [173] G. Valenza, A. Lanata, and E. P. Scilingo, "The role of nonlinear dynamics in affective valence and arousal recognition," *IEEE Trans. on Affective Computing*, vol. 3, no. 2, pp. 237–249, 2011.
- [174] W. M. Wong, A. W. Tan, C. K. Loo, and W. S. Liew, "PSO optimization of synergetic neural classifier for multichannel emotion recognition," in *World Congress on Nature and Biologically Inspired Computing*, pp. 316–321, IEEE, 2010.
- [175] L. Mirmohamadsadeghi, A. Yazdani, and J.-M. Vesin, "Using cardio-respiratory signals to recognize emotions elicited by watching music video clips," in *IEEE Int'l Workshop on Multimedia Signal Processing*, pp. 1–5, IEEE, 2016.
- [176] C.-K. Wu, P.-C. Chung, and C.-J. Wang, "Representative segment-based emotion analysis and classification with automatic respiration signal segmentation," *IEEE Trans. on Affective Computing*, vol. 3, no. 4, pp. 482–495, 2012.
- [177] X. Li, D. Song, P. Zhang, G. Yu, Y. Hou, and B. Hu, "Emotion recognition from multi-channel EEG data through convolutional recurrent neural network," in *IEEE Prof. of the Int'l Conf. on bioinformatics and biomedicine*, pp. 352–359, IEEE, 2016.
- [178] Z. Guendil, Z. Lachiri, C. Maaoui, and A. Pruski, "Emotion recognition from physiological signals using fusion of wavelet based features," in *Proc. of the Int'l Conf. on Modelling, Identification and Control*, pp. 1–6, IEEE, 2015.
- [179] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, and J.-H. Chen, "EEG-based emotion recognition in music listening," *IEEE Trans. on Biomedical Engineering*, vol. 57, no. 7, pp. 1798–1806, 2010.
- [180] B. Cheng and G. Liu, "Emotion recognition from surface EMG signal using wavelet transform and neural network," in *Prof. of the Int'l Conf. on Bioinformatics and Biomedical Engineering*, pp. 1363–1366, IEEE, 2008.

- [181] S. Basu, N. Jana, A. Bag, M. Mahadevappa, J. Mukherjee, S. Kumar, and R. Guha, "Emotion recognition based on physiological signals using valence-arousal model," in *Proc. of the Int'l Conf. on Image Information Processing*, pp. 50–55, IEEE, 2015.
- [182] J. Liu, H. Meng, A. Nandi, and M. Li, "Emotion detection from EEG recordings," in *Proc. of the Int'l Conf. on natural computation, fuzzy systems and knowledge discovery*, pp. 1722–1727, IEEE, 2016.
- [183] M. Monajati, S. H. Abbasi, F. Shabaninia, and S. Shamekhi, "Emotions states recognition based on physiological parameters by employing of fuzzy-adaptive resonance theory," 2012.
- [184] Z. Lan, O. Sourina, L. Wang, and Y. Liu, "Real-time EEG-based emotion monitoring using stable features," *The Visual Computer*, vol. 32, pp. 347–358, 2016.
- [185] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, "Identifying stable patterns over time for emotion recognition from EEG," *IEEE Trans. on Affective Computing*, vol. 10, no. 3, pp. 417–429, 2017.
- [186] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Trans. on pattern analysis and machine intelligence*, vol. 23, no. 10, pp. 1175–1191, 2001.
- [187] A. Haag, S. Goronzy, P. Schaich, and J. Williams, "Emotion recognition using bio-sensors: First steps towards an automatic system," in *Tutorial and research workshop on affective dialogue systems*, pp. 36–48, Springer, 2004.
- [188] C. L. Lisetti and F. Nasoz, "Using noninvasive wearable computers to recognize human emotions from physiological signals," *EURASIP Journal on Advances in Signal Processing*, vol. 2004, pp. 1–16, 2004.
- [189] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. on intelligent transportation systems*, vol. 6, no. 2, pp. 156–166, 2005.
- [190] E. Leon, G. Clarke, V. Callaghan, and F. Sepulveda, "A user-independent real-time emotion recognition system for software agents in domestic environments," *Engineering applications of artificial intelligence*, vol. 20, no. 3, pp. 337–345, 2007.
- [191] J. Zhai and A. Barreto, "Stress detection in computer users through non-invasive monitoring of physiological signals," *Blood*, vol. 5, no. 0, 2008.
- [192] C. D. Katsis, N. Katertsidis, G. Ganiatsas, and D. I. Fotiadis, "Toward emotion recognition in car-racing drivers: A biosignal processing approach," *IEEE Trans. on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 38, no. 3, pp. 502–512, 2008.
- [193] R. A. Calvo, I. Brown, and S. Scheduling, "Effect of experimental factors on the recognition of affective mental states through physiological measures," in *AI 2009: Advances in Artificial Intelligence: Australasian Joint Conference, Melbourne, Australia*, pp. 62–70, Springer, 2009.
- [194] G. Chanel, J. J. Kierkels, M. Soleymani, and T. Pun, "Short-term emotion assessment in a recall paradigm," *Int'l Journal of Human-Computer Studies*, vol. 67, no. 8, pp. 607–627, 2009.
- [195] Z. Khalili and M. H. Moradi, "Emotion recognition system using brain and peripheral signals: using correlation dimension to improve the results of EEG," in *Int'l Joint Conf. on Neural Networks*, pp. 1571–1575, IEEE, 2009.
- [196] J. Healey, L. Nachman, S. Subramanian, J. Shahabdeen, and M. Morris, "Out of the lab and into the fray: Towards modeling emotion in everyday life," in *Proc. of the Int'l Conf. Pervasive Computing*, pp. 156–173, Springer, 2010.
- [197] K. Plarre, A. Raji, S. M. Hossain, A. A. Ali, M. Nakajima, M. Al'Absi, E. Ertin, T. Kamarck, S. Kumar, M. Scott, *et al.*, "Continuous inference of psychological stress from sensory measurements collected in the natural environment," in *Proc. of the ACM/IEEE Int'l Conf. on information processing in sensor networks*, pp. 97–108, IEEE, 2011.
- [198] J. Hernandez, R. R. Morris, and R. W. Picard, "Call center stress recognition with person-specific models," in *Proc. of the Int'l Conf. on Affective Computing and Intelligent Interaction*, pp. 125–134, Springer, 2011.

- [199] G. Valenza, L. Citi, A. Lanatá, E. P. Scilingo, and R. Barbieri, "Revealing real-time emotional responses: a personalized assessment based on heartbeat dynamics," *Scientific Reports*, vol. 4, no. 1, p. 4998, 2014.
- [200] A. Sano and R. W. Picard, "Stress recognition using wearable sensors and mobile phones," in *Humaine association Conf. on affective computing and intelligent interaction*, pp. 671–676, IEEE, 2013.
- [201] H. P. Martinez, Y. Bengio, and G. N. Yannakakis, "Learning deep physiological models of affect," *IEEE Computational intelligence magazine*, vol. 8, no. 2, pp. 20–33, 2013.
- [202] P. Adams, M. Rabbi, T. Rahman, M. Matthews, A. Voids, G. Gay, T. Choudhury, and S. Voids, "Towards personal stress informatics: Comparing minimally invasive techniques for measuring daily stress in the wild," in *Proc. of the Int'l Conf. on Pervasive Computing Technologies for Healthcare*, pp. 72–79, 2014.
- [203] K. Hovsepian, M. Al'Absi, E. Ertin, T. Kamarck, M. Nakajima, and S. Kumar, "cStress: towards a gold standard for continuous stress assessment in the mobile environment," in *Proc. of the 2015 ACM Int'l joint conference on pervasive and ubiquitous computing*, pp. 493–504, 2015.
- [204] J. Rubin, R. Abreu, S. Ahern, H. Eldardiry, and D. G. Bobrow, "Time, frequency & complexity analysis for recognizing panic states from physiologic time-series," in *PervasiveHealth*, pp. 81–88, 2016.
- [205] N. Jaques, S. Taylor, E. Nosakhare, A. Sano, and R. Picard, "Multi-task learning for predicting health, stress, and happiness," in *NIPS Workshop on Machine Learning for Healthcare*, 2016.
- [206] A. Zenonos, A. Khan, G. Kalogridis, S. Vatsikas, T. Lewis, and M. Sooriyabandara, "Healthyoffice: Mood recognition at work using smartphones and wearable sensors," in *IEEE Int'l Conf. on Pervasive Computing and Communication Workshops*, pp. 1–6, IEEE, 2016.
- [207] M. Gjoreski, M. Luštrek, M. Gams, and H. Gjoreski, "Monitoring stress with a wrist device using context," *Journal of biomedical informatics*, vol. 73, pp. 159–170, 2017.
- [208] O. M. Mozos, V. Sandulescu, S. Andrews, D. Ellis, N. Bellotto, R. Dobrescu, and J. M. Ferrandez, "Stress detection using wearable physiological and sociometric sensors," *Int'l journal of neural systems*, vol. 27, no. 02, p. 1650041, 2017.
- [209] H. Tang, W. Liu, W.-L. Zheng, and B.-L. Lu, "Multimodal emotion recognition using deep neural networks," in *Proc. of the Int'l Conf. Neural Information Processing*, pp. 811–819, Springer, 2017.
- [210] W. Liu, W.-L. Zheng, and B.-L. Lu, "Emotion recognition using multimodal deep learning," in *Proc. of the Int'l Conf. on Neural Information Processing*, pp. 521–529, Springer, 2016.
- [211] S. Tripathi, S. Acharya, R. Sharma, S. Mittal, and S. Bhattacharya, "Using deep and convolutional neural networks for accurate emotion classification on deep data," in *Proc. of the AAAI Conf. on Artificial Intelligence*, vol. 31, pp. 4746–4752, 2017.
- [212] S. Zhao, G. Ding, J. Han, and Y. Gao, "Personality-aware personalized emotion recognition from physiological signals," in *IJCAI*, pp. 1660–1667, 2018.
- [213] L. Santamaria-Granados, M. Munoz-Organero, G. Ramirez-Gonzalez, E. Abdulhay, and N. Arunkumar, "Using deep convolutional neural network for emotion detection on a physiological signals dataset (AMIGOS)," *IEEE Access*, vol. 7, pp. 57–67, 2018.
- [214] J. Lee and S. K. Yoo, "Design of user-customized negative emotion classifier based on feature selection using physiological signal sensors," *Sensors*, vol. 18, no. 12, p. 4253, 2018.
- [215] W. Yang, M. Rifqi, C. Marsala, and A. Pinna, "Physiological-based emotion detection and recognition in a video game context," in *Proc. of the Int'l joint conf. on neural networks*, pp. 1–8, IEEE, 2018.
- [216] W. Lin, C. Li, and S. Sun, "Deep convolutional neural network for emotion recognition using EEG and peripheral physiological signal," in *Proc. of the Int'l Conf. Image and Graphics*, pp. 385–394, Springer, 2017.

- [217] A. Anusha, J. Jose, S. Preejith, J. Jayaraj, and S. Mohanasankar, "Physiological signal based work stress detection using unobtrusive sensors," *Biomedical Physics & Engineering Express*, vol. 4, no. 6, p. 065001, 2018.
- [218] S. Devi and S. Nandyala, "Electroencephalography and physiological signals for emotion analysis," *Int. Journal Innov. Technol. Exploring Eng.*, vol. 8, pp. 293–297, 2019.
- [219] L. Xia, A. S. Malik, and A. R. Subhani, "A physiological signal-based method for early mental-stress detection," in *Cyber-Enabled Intelligence*, pp. 259–289, Taylor & Francis, 2019.
- [220] H.-W. Guo, Y.-S. Huang, C.-H. Lin, J.-C. Chien, K. Haraikawa, and J.-S. Shieh, "Heart rate variability signal features for emotion recognition by using principal component analysis and support vectors machine," in *IEEE Proc. of the Int'l Conf. on bioinformatics and bioengineering*, pp. 274–277, IEEE, 2016.
- [221] H. F. García, M. A. Álvarez, and Á. Á. Orozco, "Gaussian process dynamical models for multimodal affect recognition," in *Annual Int'l Conf. of the IEEE Engineering in Medicine and Biology Society*, pp. 850–853, IEEE, 2016.
- [222] P. Rani, C. Liu, N. Sarkar, and E. Vanman, "An empirical study of machine learning techniques for affect recognition in human–robot interaction," *Pattern Analysis and Applications*, vol. 9, pp. 58–69, 2006.
- [223] J. Wagner, J. Kim, and E. André, "From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification," in *IEEE Int'l Conf. on multimedia and expo*, pp. 940–943, IEEE, 2005.
- [224] Z. Zhu, H. F. Satizabal, U. Blanke, A. Perez-Urbe, and G. Tröster, "Naturalistic recognition of activities and mood using wearable electronics," *IEEE Trans. on Affective Computing*, vol. 7, no. 3, pp. 272–285, 2015.
- [225] J. Birjandtalab, D. Cogan, M. B. Pouyan, and M. Nourani, "A non-EEG biosignals dataset for assessment and visualization of neurological status," in *IEEE Int'l Workshop on Signal Processing Systems*, pp. 110–114, IEEE, 2016.
- [226] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, 2006.
- [227] X. Zhang, C. Xu, W. Xue, J. Hu, Y. He, and M. Gao, "Emotion recognition based on multichannel physiological signals with comprehensive nonlinear processing," *Sensors*, vol. 18, no. 11, p. 3886, 2018.
- [228] D. Ayata, Y. Yaslan, and M. E. Kamasak, "Emotion recognition from multimodal physiological signals for emotion aware healthcare systems," *Journal of Medical and Biological Engineering*, vol. 40, pp. 149–157, 2020.
- [229] D. Batista, H. da Silva, A. L. N. Fred, C. F. Moreira, and H. F. Ferreira, "Benchmarking of the BITalino biomedical toolkit against an established gold standard," *Healthcare Technology Letters*, vol. 6, no. 2, pp. 32–36, 2019.
- [230] P. E. Hart, D. G. Stork, and R. O. Duda, *Pattern classification*. Wiley Hoboken, 2000.
- [231] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [232] E. Sundstrom and I. Altman, "Interpersonal relationships and personal space: Research review and theoretical model," *Human Ecology*, vol. 4, no. 1, pp. 47–67, 1976.
- [233] C. Von Scheve and S. Ismer, "Towards a theory of collective emotions," *Emotion review*, vol. 5, no. 4, pp. 406–413, 2013.
- [234] A. Dhall, A. Kaur, R. Goecke, and T. Gedeon, "EmotiW 2018: Audio-video, student engagement and group-level affect prediction," 2018.
- [235] A. Dhall, G. Sharma, R. Goecke, and T. Gedeon, "EmotiW 2020: Driver gaze, group emotion, student engagement and physiological signal based challenges," in *Proc. of the Int'l Conf. on Multimodal Interaction*, pp. 784–789, ACM, 2020.
- [236] M. Iwasaki and Y. Noguchi, "Hiding true emotions: Micro-expressions in eyes retrospectively concealed by mouth movements," *Scientific Reports*, vol. 6, no. 1, p. 22049, 2016.

- [237] C. Herrando and E. Constantinides, "Emotional contagion: A brief overview and future directions," *Frontiers in Psychology*, vol. 12, p. 712606, 2021.
- [238] A. Strang, G. Funke, S. Russell, A. Dukes, and M. Middendorf, "Physio-behavioral coupling in a cooperative team task: Contributors and relations," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 40, no. 1, p. 145, 2014.
- [239] E. Delaherche, M. Chetouani, A. Mahdhaoui, C. Saint-Georges, S. Viaux, and D. Cohen, "Interpersonal synchrony: a survey of evaluation methods across disciplines," *IEEE Trans. on Affective Computing*, vol. 3, no. 3, pp. 349–365, 2012.
- [240] V. Misal, S. Akiri, S. Taherzadeh, H. McGowan, G. Williams, J. L. Jenkins, H. Mentis, and A. Kleinsmith, "Physiological synchrony, stress and communication of paramedic trainees during emergency response training," in *Companion Publication of the Int'l Conf. on Multimodal Interaction*, p. 82–86, ACM, 2020.
- [241] A. Karvonen, V.-L. Kykyri, J. Kaartinen, M. Penttonen, and J. Seikkula, "Sympathetic nervous system synchrony in couple therapy," *Journal Marital Fam. Ther.*, vol. 42, no. 3, pp. 383–395, 2016.
- [242] W.-S. Chien, H.-C. Chou, and C.-C. Lee, "Self-assessed emotion classification from acoustic and physiological features within small-group conversation," *Companion Publication of the Int'l Conf. on Multimodal Interaction*, pp. 230–239, 2021.
- [243] S. Mariooryad and C. Busso, "Exploring cross-modality affective reactions for audiovisual emotion recognition," *IEEE Trans. on Affective Computing*, vol. 4, no. 2, pp. 183–196, 2013.
- [244] X. Guo, L. Polanía, and K. Barner, "Group-level emotion recognition using deep models on image scene, faces, and skeletons," in *Proc. of the ACM Int'l. Conf. on Multimodal Interaction*, p. 603–608, ACM, 2017.
- [245] B. Nagarajan and R. Oruganti, "Group emotion recognition in adverse face detection," in *IEEE Int'l Conf. on Automatic Face Gesture Recognition*, pp. 1–5, 2019.
- [246] J. Quan, Y. Miyake, and T. Nozawa, "Incorporating interpersonal synchronization features for automatic emotion recognition from visual and audio data during communication," *Sensors*, vol. 21, no. 16, p. 5317, 2021.
- [247] G. Chanel, S. Avry, G. Molinari, M. Bétrancourt, and T. Pun, "Multiple users' emotion recognition: Improving performance by joint modeling of affective reactions," in *Int'l Conf. on Affective Computing and Intelligent Interaction*, pp. 92–97, 2017.
- [248] W.-S. Chien, H.-C. Chou, and C.-C. Lee, "Belongingness and satisfaction recognition from physiological synchrony with a group-modulated attentive BLSTM under small-group conversation," *Companion Publication of the Int'l Conf. on Multimodal Interaction*, pp. 220–229, 2021.
- [249] R. Palumbo, M. Marraccini, L. Weyandt, O. Wilder-Smith, H. McGee, S. Liu, and M. Goodwin, "Interpersonal autonomic physiology: A systematic review of the literature," *Personality and Social Psychology Review*, vol. 21, no. 2, pp. 99–141, 2017.
- [250] G. Chanel, M. Kivikangas, and N. Ravaja, "Physiological compliance for social gaming analysis: Cooperative versus competitive play," *Interacting with Computers*, vol. 24, no. 4, pp. 306–316, 2012.
- [251] S. Järvelä, J. M. Kivikangas, J. Kätsyri, and N. Ravaja, "Physiological linkage of dyadic gaming experience," *Simulation & Gaming*, vol. 45, no. 1, pp. 24–40, 2014.
- [252] R. Reed, A. Randall, J. Post, and E. Butler, "Partner influence and in-phase versus anti-phase physiological linkage in romantic couples," *Int'l. Journal of Psychophysiology*, vol. 88, no. 3, pp. 309–316, 2013.
- [253] R. Silver and R. Parente, "The psychological and physiological dynamics of a simple conversation," *Social Behavior and Personality: An Int'l Journal*, vol. 32, no. 5, pp. 413–418, 2004.
- [254] D. Shearn, L. Spellman, B. Straley, J. Meirick, and K. Stryker, "Empathic blushing in friends and strangers," *Motivation and Emotion*, vol. 23, no. 4, pp. 307–316, 1999.

- [255] I. Konvalinka, D. Xygalatas, J. Bulbulia, U. Schjødt, E.-M. Jegindø, S. Wallot, G. Orden, and A. Roepstorff, "Synchronized arousal between performers and related spectators in a fire-walking ritual," *Proc. of the National Academy of Sciences*, vol. 108, no. 20, pp. 8514–8519, 2011.
- [256] P. Mitkidis, J. McGraw, A. Roepstorff, and S. Wallot, "Building trust: Heart rate synchrony and arousal during joint action increased by public goods game," *Physiology & Behavior*, vol. 149, pp. 101–106, 2015.
- [257] E. Montague, J. Xu, and E. Chiou, "Shared experiences of technology and trust: An experimental study of physiological compliance between active and passive users in technology-mediated collaborative encounters," *IEEE Trans. on Human-Machine Systems*, vol. 44, no. 5, pp. 614–624, 2014.
- [258] V. Müller and U. Lindenberger, "Cardiac and respiratory patterns synchronize between persons during choir singing," *PloS one*, vol. 6, no. 9, pp. 1–15, 2011.
- [259] E. Ferrer and J. L. Helm, "Dynamical systems modeling of physiological coregulation in dyadic interactions," *Int'l. Journal of Psychophysiology*, vol. 88, no. 3, pp. 296–308, 2013.
- [260] A. Bachrach, Y. Fontbonne, C. Joufflineau, and J. L. Ulloa, "Audience entrainment during live contemporary dance performance: physiological and cognitive measures," *Frontiers in Human Neuroscience*, vol. 9, p. 179, 2015.
- [261] E. Codrons, N. Bernardi, M. Vandoni, and L. Bernardi, "Spontaneous group synchronization of movements and respiratory rhythms," *PloS one*, vol. 9, no. 9, pp. 1–10, 2014.
- [262] L. Nummenmaa, J. M. Lahnakoski, and E. Glerean, "Sharing the social world via intersubject neural synchronisation," *Current Opinion in Psychology*, vol. 24, pp. 7–14, 2018. Social Neuroscience.
- [263] S. Järvelä, *Physiological Synchrony and Affective Protosocial Dynamics*. PhD thesis, University of Helsinki Faculty of Arts, 2020.
- [264] E. Jun, D. McDuff, and M. Czerwinski, "Circadian rhythms and physiological synchrony: Evidence of the impact of diversity on small group creativity," *Proc. ACM Hum. Comput. Interact.*, vol. 3, pp. 1–22, 2019.
- [265] Z. Li, M. Sturge-Apple, S. Liu, and P. Davies, "Parent-adolescent physiological synchrony: Moderating effects of adolescent emotional insecurity," *Psychophysiology*, vol. 57, no. 9, p. e13596, 2020.
- [266] E. Prochazkova, E. Sjak-Shie, F. Behrens, D. Lindh, and M. Kret, "Physiological synchrony is associated with attraction in a blind date setting," *Nature Human Behaviour*, vol. 6, no. 2, pp. 269–278, 2022.
- [267] S. Gashi, E. Di Lascio, and S. Santini, "Using students' physiological synchrony to quantify the classroom emotional climate," in *Proc. Int'l Joint Conf. and Int'l Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, p. 698–701, ACM, 2018.
- [268] I. Gordon, S. Wallot, and Y. Berson, "Group-level physiological synchrony and individual-level anxiety predict positive affective behaviors during a group decision-making task," *Psychophysiology*, vol. 58, no. 9, p. e13857, 2021.
- [269] D. Richardson and R. Dale, "Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension," *Cognitive Science*, vol. 29, no. 6, pp. 1045–1060, 2005.
- [270] A. Knight, D. Kennedy, and S. McComb, "Using recurrence analysis to examine group dynamics," *Group dynamics: Theory, Research, and Practice*, vol. 20, no. 3, p. 223, 2016.
- [271] A. Guidi, A. Lanata, P. Baragli, G. Valenza, and E. Scilingo, "A wearable system for the evaluation of the human-horse interaction: A preliminary study," *Electronics*, vol. 5, no. 4, 2016.
- [272] P. Chikersal, M. Tomprou, Y. Kim, A. Woolley, and L. Dabbish, "Deep structures of collaboration: Physiological correlates of collective intelligence and group satisfaction," in *Proc. of the ACM Conf. on Computer Supported Cooperative Work and Social Computing*, p. 873–888, ACM, 2017.

- [273] O. Oullier, G. Guzman, K. Jantzen, J. Lagarde, and J. Kelso, "Social coordination dynamics: Measuring human bonding," *Social Neuroscience*, vol. 3, no. 2, pp. 178–192, 2008.
- [274] E. Delaherche and M. Chetouani, "Multimodal coordination: Exploring relevant features and measures," in *Proc. of the Int'l Workshop on Social Signal Processing*, p. 47–52, ACM, 2010.
- [275] P. Hamilton, "Open source ECG analysis," in *Computers in Cardiology*, pp. 101–104, 2002.
- [276] M. Elgendi, I. Norton, M. Brearley, D. Abbott, and D. Schuurmans, "Systolic peak detection in acceleration photoplethysmograms measured from emergency responders in tropical conditions," *PloS one*, vol. 8, no. 10, pp. 1–11, 2013.
- [277] T. Zhang, A. E. Ali, C. Wang, A. Hanjalic, and P. Cesar, "Weakly-supervised learning for fine-grained emotion recognition using physiological signals," *IEEE Trans. on Affective Computing*, 2022.
- [278] S. Jerritta, M. Murugappan, R. Nagarajan, and K. Wan, "Physiological signals based human emotion recognition: A review," *Proc. of the IEEE Int'l Colloquium on Signal Processing and its Applications*, vol. 1, pp. 410–415, 2011.
- [279] A. Greco, G. Valenza, A. Lanata, E. Scilingo, and L. Citi, "cvxEDA: A convex optimization approach to electrodermal activity processing," *IEEE Trans. on Biomedical Engineering*, vol. 63, no. 4, pp. 797–804, 2016.
- [280] R. Elalamy, M. Fanourakis, and G. Chanel, "Multi-modal emotion recognition using recurrence plots and transfer learning on physiological signals," in *Proc. of the IEEE Int'l Conf. on Affective Computing and Intelligent Interaction*, pp. 1–7, 2021.
- [281] Siddharth, T. Jung, and T. Sejnowski, "Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing," *IEEE Trans. on Affective Computing*, vol. 13, no. 1, pp. 96–107, 2022.
- [282] J. Shukla, M. Barreda-Ángeles, J. Oliver, G. Nandi, and D. Puig, "Feature extraction and selection for emotion recognition from electrodermal activity," *IEEE Trans. on Affective Computing*, vol. 12, no. 4, pp. 857–869, 2021.
- [283] G. Yin, S. Sun, D. Yu, D. Li, and K. Zhang, "A multimodal framework for large-scale emotion recognition by fusing music and electrodermal activity signals," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 18, no. 3, 2022.
- [284] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.
- [285] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, pp. 8024–8035, Curran Associates, Inc., 2019.
- [286] R. Liaw, E. Liang, R. Nishihara, P. Moritz, J. E. Gonzalez, and I. Stoica, "Tune: A research platform for distributed model selection and training," *arXiv preprint arXiv:1807.05118*, 2018.
- [287] P. Gupta, S. A. Balaji, S. Jain, and R. K. Yadav, "Emotion recognition during social interactions using peripheral physiological signals," in *Computer Networks and Inventive Communication Technologies*, pp. 99–112, Springer Singapore, 2022.
- [288] F. Ramseyer and W. Tschacher, *Nonverbal Synchrony or Random Coincidence? How to Tell the Difference*, pp. 182–196. Springer Berlin Heidelberg, 2010.
- [289] R. G. Moulder, S. M. Boker, F. Ramseyer, and W. Tschacher, "Determining synchrony between behavioral time series: An application of surrogate data generation for establishing falsifiable null-hypotheses.," *Psychological Methods*, vol. 23, no. 4, p. 757, 2018.
- [290] S. M. Montgomery, N. Nair, P. Chen, and S. Dikker, "Introducing emotibit, an open-source multi-modal sensor for measuring research-grade physiological signals," *Science Talks*, vol. 6, p. 100181, 2023.
- [291] M. Abreu, A. Fred, H. Plácido Da Silva, and C. Wang, "From seizure detection to prediction: A review of wearables and related devices applicable to epilepsy via peripheral measurements," tech. rep., Institute of Telecommunications, 2020.

- [292] M. de Zambotti, L. Rosas, I. M. Colrain, and F. C. Baker, "The sleep of the ring: Comparison of the ōura sleep tracker against polysomnography," *Behavioral Sleep Medicine*, vol. 17, no. 2, pp. 124–136, 2019.
- [293] S. Kwon, J. Hong, E.-K. Choi, B. Lee, C. Baik, E. Lee, E.-R. Jeong, B.-K. Koo, S. Oh, and Y. Yi, "Detection of atrial fibrillation using a ring-type wearable device (cardiotracker) and deep learning analysis of photoplethysmography signals: Prospective observational proof-of-concept study," *Journal Med. Internet Res.*, vol. 22, no. 5, p. e16443, 2020.
- [294] J. E. Collins, H. Y. Yang, T. P. Trentadue, Y. Gong, and E. Losina, "Validation of the Fitbit charge 2 compared to the actigraph GT3X+ in older adults with knee osteoarthritis in free-living conditions," *PloS one*, vol. 14, no. 1, pp. 1–14, 2019.
- [295] M. Grinberg, *Flask web development: developing web applications with python*. " O'Reilly Media, Inc.", 2018.
- [296] The HDF Group, "Hierarchical Data Format, version 5."
- [297] G. N. Yannakakis and H. P. Martínez, "Grounding truth via ordinal annotation," in *Proc. of the Int'l Conf. on Affective Computing and Intelligent Interaction*, pp. 574–580, 2015.
- [298] T. Zhang, A. El Ali, C. Wang, A. Hanjalic, and P. Cesar, *RCEA: Real-Time, Continuous Emotion Annotation for Collecting Precise Mobile Video Ground Truth Labels*, p. 1–15. ACM, 2020.
- [299] J. M. Girard and A. G. C. Wright, "DARMA: Software for dual axis rating and media annotation," *Behavior Research Methods*, vol. 50, no. 3, pp. 902–909, 2018.
- [300] D. Melhart, A. Liapis, and G. N. Yannakakis, "PAGAN: Video affect annotation made easy," 2019.
- [301] J. Broekens and W.-P. Brinkman, "AffectButton: A method for reliable and valid affective self-report," *Int'l Journal of Human-Computer Studies*, vol. 71, no. 6, pp. 641–667, 2013.
- [302] P. Lopes, G. N. Yannakakis, and A. Liapis, "RankTrace: Relative and unbounded affect annotation," in *Proc. of the Int'l Conf. on Affective Computing and Intelligent Interaction*, pp. 158–163, 2017.
- [303] T. Baur, A. Heimerl, F. Lingenfelser, J. Wagner, M. F. Valstar, B. Schuller, and E. André, "eXplainable cooperative machine learning with NOVA," *KI-Künstliche Intelligenz*, vol. 34, pp. 143–164, 2020.
- [304] N. Runge, M. Hellmeier, D. Wenig, and R. Malaka, "Tag your emotions: A novel mobile user interface for annotating images with emotions," in *Proc. of the Int'l Conf. on Human-Computer Interaction with Mobile Devices and Services Adjunct*, p. 846–853, ACM, 2016.
- [305] E. Gatti, E. Calzolari, E. Maggioni, and M. Obrist, "Emotional ratings and skin conductance response to visual, auditory and haptic stimuli," *Scientific Data*, vol. 5, no. 1, p. 180120, 2018.
- [306] T. Schmidt, I. Engl, D. Halhuber, and C. Wolff, "Comparing live sentiment annotation of movies via arduino and a slider with textual annotation of subtitles," in *Post-Proc. of the Conf. Digital Humanities in the Nordic Countries*, vol. 2865, pp. 212–223, CEUR Workshop Proc., 2021.
- [307] M. Robinson and G. Clore, "Belief and feeling: Evidence for an accessibility model of emotional self-report," *Psychological Bulletin*, vol. 128, pp. 934–60, 2002.
- [308] L. F. Barrett, "The relationships among momentary emotion experiences, personality descriptions, and retrospective ratings of emotion," *Personality and Social Psychology Bulletin*, vol. 23, no. 10, pp. 1100–1110, 1997.
- [309] C. Röcke, C. A. Hoppmann, and P. L. Klumb, "Correspondence Between Retrospective and Momentary Ratings of Positive and Negative Affect in Old Age: Findings From a One-Year Measurement Burst Design," *The Journals of Gerontology: Series B*, vol. 66B, no. 4, pp. 411–415, 2011.
- [310] E. Öhman, "Challenges in annotation: annotator experiences from a crowdsourced emotion annotation task," in *Proc. of the Digital Humanities in the Nordic Countries Conference*, pp. 293 – 301, CEUR Workshop Proc., 2020.
- [311] I. Siegert, R. Böck, and A. Wendemuth, "Inter-rater reliability for emotion annotation in human-computer interaction: comparison and methodological improvements," *Journal on Multimodal User Interfaces*, vol. 8, no. 1, pp. 17–28, 2014.

- [312] A. Mill, A. Realo, and J. Allik, "Retrospective ratings of emotions: the effects of age, daily tiredness, and personality," *Frontiers in Psychology*, vol. 6, 2016.
- [313] R. R. McCrae and O. P. John, "An introduction to the five-factor model and its applications," *Journal of Personality*, vol. 60, no. 2, pp. 175–215, 1992.
- [314] H. L. Urry and J. J. Gross, "Emotion regulation in older age," *Current Directions in Psychological Science*, vol. 19, no. 6, pp. 352–357, 2010.
- [315] E. Schryer and M. Ross, "Evaluating the valence of remembered events: The importance of age and self-relevance," *Psychology and Aging*, vol. 27, pp. 237–42, 2011.
- [316] D. L. Paulhus and O. P. John, "Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives," *Journal of Personality*, vol. 66, no. 6, pp. 1025–1060, 1998.
- [317] R. D. Lane, G. L. Ahern, G. E. Schwartz, and A. W. Kaszniak, "Is alexithymia the emotional equivalent of blindsight?," *Biological Psychiatry*, vol. 42, no. 9, pp. 834–844, 1997.
- [318] W. Boucsein, *Methods of Electrodermal Recording*, pp. 87–258. Springer US, 2012.
- [319] C. Liu, A. Shmilovici, and M. Last, "Towards story-based classification of movie scenes," *PloS one*, vol. 15, no. 2, pp. 1–22, 2020.
- [320] M. Mather, "Emotional memory," *The Encyclopedia of Adulthood and Aging*, pp. 1–4, 2015.
- [321] J. Brooke, "SUS: a retrospective," *Journal of Usability Studies Archive*, vol. 8, pp. 29–40, 2013.
- [322] R. A. Grier, "How high is high? a meta-analysis of NASA-TLX global workload scores," *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, vol. 59, no. 1, pp. 1727–1731, 2015.
- [323] J. R. Lewis and J. Sauro, "Item benchmarks for the system usability scale," *Journal Usability Studies*, vol. 13, no. 3, p. 158–167, 2018.
- [324] M. Grimm and K. Kroschel, "Evaluation of natural emotions using self assessment manikins," in *Proc. of the IEEE Workshop on Automatic Speech Recognition and Understanding*, pp. 381–385, 2005.
- [325] N. Malandrakis, A. Potamianos, G. Evangelopoulos, and A. Zlatintsi, "A supervised approach to movie emotion tracking," in *Proc. of the IEEE Int'l Conf. on Acoustics, Speech and Signal Processing*, pp. 2376–2379, 2011.
- [326] T. Li, Y. Baveye, C. Chamaret, E. Dellandréa, and L. Chen, "Continuous arousal self-assessments validation using real-time physiological responses," in *Proc. of the Int'l Workshop on Affect & Sentiment in Multimedia*, p. 39–44, ACM, 2015.
- [327] J. Fleureau, P. Guillotel, and I. Orlac, "Affective benchmarking of movies based on the physiological responses of a real audience," in *Proc. of the Humaine Association Conf. on Affective Computing and Intelligent Interaction*, pp. 73–78, 2013.
- [328] R. Cowie, G. McKeown, and E. Douglas-Cowie, "Tracing emotion: an overview," *Int'l Journal of Synthetic Emotions*, vol. 3, no. 1, pp. 1–17, 2012.
- [329] C. C. Musat, A. Ghasemi, and B. Faltings, "Sentiment analysis using a novel human computation game," in *Proc. of the Workshop on the People's Web Meets NLP: Collaboratively Constructed Semantic Resources and their Applications to NLP*, pp. 1–9, Association for Computational Linguistics, 2012.
- [330] A. Metallinou and S. Narayanan, "Annotation and processing of continuous emotional attributes: Challenges and opportunities," in *IEEE Int'l Conf. and Workshops on Automatic Face and Gesture Recognition*, pp. 1–8, 2013.
- [331] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A database for facial expression, valence, and arousal computing in the wild," *IEEE Trans. on Affective Computing*, vol. 10, no. 01, pp. 18–31, 2019.
- [332] A. Dhali, R. Goecke, J. Joshi, M. Wagner, and T. Gedeon, "Emotion recognition in the wild challenge (EmotiW) challenge and workshop summary," in *Proc. of the Int'l Conf. on Multimodal Interaction*, pp. 371–372, 2013.

- [333] S. Saganowski, J. Komoszyńska, M. Behnke, B. Perz, D. Kunc, B. Klich, Ł. D. Kaczmarek, and P. Kazienko, “Emognition dataset: emotion recognition with self-reports, facial expressions, and physiology using wearables,” *Scientific data*, vol. 9, no. 1, p. 158, 2022.
- [334] M. Behnke, M. Buchwald, A. Bykowski, S. Kupiński, M. Kosakowski, J. Enko, D. Drażkowski, and L. D. Kaczmarek, “POPANE dataset - psychophysiology of positive and negative emotions.” 10.17605/OSF.IO/94BPX (2023).
- [335] W. Li, R. Tan, Y. Xing, G. Li, S. Li, G. Zeng, P. Wang, B. Zhang, X. Su, D. Pi, G. Guo, and D. Cao, “A multimodal psychological, physiological and behavioural dataset for human emotions in driving tasks,” vol. 9, no. 1, p. 481. figshare 10.6084/m9.figshare.c.5744171.v1 (2022).
- [336] M. Zhang, L. Yu, K. Zhang, B. Du, B. Zhan, S. Chen, X. Jiang, S. Guo, J. Zhao, Y. Wang, B. Wang, S. Liu, and W. Luo, “Kinematic dataset of actors expressing emotions,” *Scientific Data*, vol. 7, no. 1, p. 292, 2020. PhysioNet <https://doi.org/10.13026/kg8b-1t49> (2020).
- [337] F. H. Wilhelm and P. Grossman, “Emotions beyond the laboratory: theoretical fundamentals, study design, and analytic strategies for advanced ambulatory assessment,” *Biological Psychology*, vol. 84, no. 3, pp. 552–569, 2010.
- [338] Y. Xu, I. Hübener, A.-K. Seipp, S. Ohly, and K. David, “From the lab to the real-world: An investigation on the influence of human movement on emotion recognition using physiological signals,” in *Proc. of the IEEE Int’l Conf. on Pervasive Computing and Communications Workshops*, pp. 345–350, 2017.
- [339] S. K. D’Mello and B. M. Booth, “Affect detection from wearables in the ‘real’ world: Fact, fantasy, or somewhere in between?,” *IEEE Intelligent Systems*, vol. 38, no. 1, pp. 76–84, 2023.
- [340] J. Russell, “Affective space is bipolar,” *Journal of Personality and Social Psychology*, vol. 37, no. 3, p. 345, 1979.
- [341] T. Canli, H. Sivers, S. Whitfield, I. Gotlib, and J. Gabrieli, “Amygdala response to happy faces as a function of extraversion,” *Science*, vol. 296, no. 5576, pp. 2191–2191, 2002.
- [342] S. B. Eysenck, H. J. Eysenck, and P. Barrett, “A revised version of the psychoticism scale,” *Personality and Individual Differences*, vol. 6, no. 1, pp. 21–29, 1985.
- [343] J. Johnson, “Measuring thirty facets of the five factor model with a 120-item public domain inventory: Development of the IPIP-NEO-120,” *Journal of Research in Personality*, vol. 51, pp. 78–89, 2014.
- [344] C. Carreiras, H. Silva, A. Lourenço, and A. Fred, “StorageBit-a metadata-aware, extensible, semantic and hierarchical database for biosignals,” in *Proc. of the Int’l Conf. on Health Informatics*, vol. 2, pp. 65–74, SciTePress, 2013.
- [345] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python,” *Nature Methods*, vol. 17, pp. 261–272, 2020.
- [346] P. W. Macfarlane, *Comprehensive electrocardiology*. New York: Springer, 2nd ed., 2011.
- [347] A. Banganho, M. Santos, and H. P. da Silva, “Electrodermal activity: Fundamental principles, measurement, and application,” *IEEE Potentials*, vol. 41, pp. 35–43, 2022.
- [348] M. Nasser, E. Nurse, M. Glasstetter, S. Böttcher, N. Gregg, A. Laks, B. Joseph, T. Pal, P. Viana, E. Bruno, A. Biondi, M. Cook, G. Worrell, A. Schulze-Bonhage, M. Dümpelmann, D. Freestone, M. Richardson, and B. Brinkmann, “Signal quality and patient experience with wearable devices for epilepsy management,” *Epilepsia*, vol. 61, no. S1, pp. S25–S35, 2020.
- [349] S. Böttcher, S. Vieluf, E. Bruno, B. Joseph, N. Epitashvili, A. Biondi, N. Zabler, M. Glasstetter, M. Dümpelmann, K. Van Laerhoven, M. Nasser, B. H. Brinkman, M. P. Richardson, A. Schulze-Bonhage, and T. Loddenkemper, “Data quality evaluation in wearable monitoring,” *Scientific Reports*, vol. 12, no. 1, p. 21412, 2022.

- [350] A. Gautam, N. Simoes-Capela, G. Schiavone, A. Acharyya, W. de Raedt, and C. Van, "A data driven empirical iterative algorithm for GSR signal pre-processing," in *European Signal Processing Conf.*, pp. 1162–1166, 2018.
- [351] M. Glasstetter, S. Böttcher, N. Zabler, N. Epitashvili, M. Dümpelmann, M. P. Richardson, and A. Schulze-Bonhage, "Identification of ictal tachycardia in focal motor- and non-motor seizures by means of a wearable PPG sensor," *Sensors*, vol. 21, no. 18, p. 6017, 2021.
- [352] J. A. Castro-García, A. J. Molina-Cantero, I. M. Gómez-González, S. Lafuente-Arroyo, and M. Merino-Monge, "Towards human stress and activity recognition: A review and a first approach based on low-cost wearables," *Electronics*, vol. 11, no. 1, p. 155, 2022.
- [353] A. Devices, "Signal-to-noise ratio as a quantitative measure for optical biosensors." <https://analog.com/en/design-notes/signal-to-noise-ratio-as-a-quantitative-measure-for-optical-biosensors.html>, 2023. Accessed: 5 July 2023.
- [354] J. Doolittle, P. Walker, T. Mills, and J. Thurston, "Hyperhidrosis: an update on prevalence and severity in the united states," *Archives of Dermatological Research*, vol. 308, no. 10, pp. 743–749, 2016.
- [355] P. Verduyn and K. Brans, "The relationship between extraversion, neuroticism and aspects of trait affect," *Personality and Individual Differences*, vol. 52, no. 6, pp. 664–669, 2012.
- [356] M. M. Bradley and P. J. Lang, "Measuring emotion: Behavior, feeling, and physiology," in *Cognitive Neuroscience of Emotion*, pp. 25–49, Oxford University Press, 2000.
- [357] J. J. Braithwaite, D. P. Z. Watson, R. O. Jones, and M. A. Rowe, "Guide for analysing electrodermal activity & skin conductance responses for psychological experiments," tech. rep., 2013.
- [358] S. Gashi, E. Di Lascio, and S. Santini, "Using unobtrusive wearable sensors to measure the physiological synchrony between presenters and audience members," *Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 1, 2019.