Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics

Rebekka Birkeland Brøyn

Emotion Detection using a Low-Cost Wearable Sensing System

Based on Physiological Indicators

Master's thesis in Cybernetics and Robotics
Supervisor: Damiano Varagnolo

February 2021

NTNU
Norwegian University of Science and Technology

Rebekka Birkeland Brøyn

Emotion Detection using a Low-Cost Wearable Sensing System

Based on Physiological Indicators

Master's thesis in Cybernetics and Robotics
Supervisor: Damiano Varagnolo
February 2021

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics

NTNU
Norwegian University of
Science and Technology

Master Thesis Description

Main objective

The main objective of the master thesis is to explore the possibilities for emotion detection based on physiological signals by using a wearable sensing system. The project concerns a wide field and the thesis will hopefully form a foundation for further research.

Tasks

- Literature study on similar experiments, examination of possible sensors to use and what possible physiological signals to monitor.
- Designing a low cost wearable sensing system.
- Designing an experimental protocol including different choices such as elicitation method of specific emotions.
- Collect data samples by conducting field experiments on volunteers according to the defined protocol.
- Analyzing the data samples and applying data driven models for detecting specific emotions.

Summary

Emotions intuitively seem like an uncontrolled mystery with an extremely large impact on our lives. Finding an explanation for different emotions has aroused the interest of many philosophers and psychologists throughout history. The field has gradually expanded and, among others, neuroscientists and engineers have joined in the quest to solve the mystery. Research has up to now shown that there exists a correlation between certain physiological signals and specific emotions, and scientists have discovered the advantage of applying data-driven models to track important indicators.

Furthermore, a huge interest has grown in monitoring personal health data to maintain a healthy lifestyle. Today, one can monitor pulse, sleep quality, activity levels, and much more with a smartwatch. There is reason to believe that the mood of a person can be estimated from such health indicators. Furthermore, the integration of sophisticated methods for directly inferring one's emotional state from physiological signals would provide exciting and useful elements to integrate smartwatches or monitoring wristbands in providing useful feedback to users themselves, and potentially their physicians.

For this purpose, we here explore the possibilities for automatically detecting and classifying human emotions by processing physiological data collected using a wearable sensing system. The thesis explores such a process from beginning to end, i.e., from the prototyping and design of the physiological signals sensing system, to organizing and conducting field experiments on volunteers by following an ad-hoc medical protocol, and finally arriving at analyzing the samples and applying machine learning algorithms for estimating and classifying the emotions of the subjects based on the well-known valence-arousal methodology.

Sammendrag

Emosjoner virker intuitivt som et unkontollert mysterie med en ekstremt stor innvirkning på livene våre. Å finne en forklaring på de ulike emosjonene har vekket interessen til mange filosofer og psykologer gjennom historien. Feltet har gradvis ekspandert og, blant annet, nevroforskere og ingenører har blitt med på oppgaven om å løse mysteriet. Forskning har opp til nå vist at det finnes sammenheng mellom visse fysiologiske signaler og spesifike emosjoner, og forskere har oppdaget fordelen med å benytte datadrevne modeller til å spore viktige indikasjoner.

Videre har vi sett en stor, økende interesse i å overvåke personlig helsedata for å ivareta en sunn livsstil. I dag kan man overvåke puls, søvnkvalitet, aktivitetslevel, og mye mer med en smartklokke. Det er grunn til å tro at humøret til et individ kan bli estimert fra slike helseindikatorer. Videre ville sofistikerte metoder for å direkte utlede ens emosjonelle tilstand fra fysiologiske signaler være et interesant og nyttig element å integrere i slike smartklokker eller eventulle overvåkende armbånd for å gi nyttig tilbakemelding til brukere eller relatert helsepersonell.

For dette formålet vil vi her undersøke mulighetene for automatisk detektering og klassifisering av menneskelige emosjoner ved å prosessere fysiologiske data, samlet ved å benytte et bærbart sensorsystem. Oppgaven undersøker en slik prosess fra begynnelse til slutt, det vil si fra prototype og design av sensorsystemet, til å organisere gjennomføre felteksperiment på frivillige ved å følge en ad-hoc medisinsk protokoll, og til slutt analysere dataprøvene og benytte makinlærings algoritmer til å estimere og klassifisere subjektens emosjoner basert på vellkjent valens-opphisselse metodikk.

Preface

This master thesis has been a part of a larger collaboration between Norwegian University of Science and Technology (NTNU) and Otto von Guericke University Magdeburg (OVGU).

First, I want to thank my supervisor, Professor Damiano Varagnolo for introducing me to a very exiting project, for being a great motivator, including me in an international research group and for his valuable guidance. In the same regard, this work could not have been done without the support of the entire team, that includes Roya Doshmanziari which has been a great support at NTNU, and Roxanne Jackson and Steffi Knorn from OVGU which has given invaluable input and guidance.

Furthermore I would like to thank Marieke Dewitte, assistant professor at the Department of Clinical Psychological Science of the University of Maastrich, The Netherlands, which has played a crucial role in the design of the experimental protocol used in the thesis. I would also like to thank Stefano Brevik Bertelli for good advancing regarding hardware and wiring.

At last, I would like to thank all the volunteers that have been participating in the experiments. This group includes friends and family that deserves a special thanks in which they have been a great support through the project and in general. The project has required some Guinea Pigs in developing a experimental protocol. Thus a big thank you and a small apology to Kari Vikøren Mo and Tiril Sundby which have been helpful in every situation.

Table of Contents

List of Tables

List of Figures

Abbreviations

ANS Autonomic Nervous System

AT ATtention

BLE Bluetooth Low Energy

BPM beats per minute

ECG Electrocardiography

EDA Electrodermal Activity

EEG Electroencephalography

EMG Electromyography

FA Film Amusement

FB Film Baseline

FF Film Fear

FFT Fast Fourier Transform

FS Film Sad

FT Film Tenderness

GDPR General Data Protection Regulation

GSR Galvanic Skin Response

HCI Human Computer Interaction

HDC Health Data Collector

HF High Frequency

HR Heart Rate

HRV Heart Rate Variability

LF Low Frequency

NSD Norwegian Center for Research Data

PC Principal Component

PCA Principal Component Analysis

PSD Power Spectral Density

PSNS Parasympathetic Nervous System

Q1 First Quadrant

Q2 Second Quadrant

Q3 Third Quadrant

Q4 Fourth Quadrant

RBF Radial Basis Fuction

SAM Self Assessment Manikin

SDK Software Development Kit

SNS Sympathetic Nervous system

SVD Singular Value Decomposition

SVM Support Vector Machines

VLF Very Low Frequency

VR Virtual Reality

Chapter 1

Introduction

1.1 Motivation

Data-driven emotion recognition systems have been well studied in the last decades and have played an important role in numerous areas. Especially is emotion recognition a prominent topic within the field of human-computer interaction in which it is an essential element in the development of affective computers. The development of affective computers requires the robot to acquire human empathetic abilities such as understanding another person's emotional state from facial expressions, body gestures, poses, or the way they speak. Recent studies have shown promising results in recognizing emotions using speech-, facial expression- and gesture recognition systems. However, it isn't always easy for a human or a computer to recognize an emotional state by evaluating the properties above since humans can also suppress indicators of emotion. Other studies suggest that by analysing physiological responses directly from the Autonomic Nervous System (ANS), suppressed emotions can be detected.

Such studies may reveal the suppressed feelings of fellow human beings, or even more importantly, better understand your own emotions. Commercial monitoring systems have become very popular in recent years. With a smartwatch, one can monitor sleep quality, activity level, pulse, and more; for the purpose of maintaining a healthy lifestyle. We pose the question: *What if we could measure happiness directly by using the same non-invasive monitoring system?* Then we could become aware of the situations that make us the happiest, seek out these situations, and avoid situations that trigger negative emotions. We should however acknowledge that long-term happiness is sometimes the result of hard work which may trigger negative emotions on short-term. Nevertheless, stress and negative emotions over time can trigger chronic diseases such as anxiety, which has become an increasing problem in society [34][32]. We believe that a smartwatch or other easily wearable devices that can detect your emotional state at any time would be an aid for emotional regulation. Thus one could maintain balance and avoid fatigue of the autonomic nervous system to help prevent chronic disorders. Furthermore, it is suggested that some

disorders reduce one's ability to regular emotions. Examples of such disorders include depression, bipolar disorder, anxiety, schizophrenia and autism [48, 11, 29, 51]. Within this topic, Garcia-Ceja et al. [19] have developed a commercial wristband that determines the different phases of bipolar disorder based on motor activity. We believe that including emotion recognition based on physiological responses will give a higher accuracy in the predictions and cover a wider range of disorders.

In this thesis we propose a low-cost, wearable sensing system and a software solution that can detect specific emotions based on physiological signals.

1.2 Literature Review

Numerous attempts have been made to detect emotions by data-driven models with a great variation in the results. These various experiments have been executed with different selections of emotional excitation methods, a variation of detectable emotions, different indicators and features used for prediction, and distinct classification algorithms. This includes different selections of emotional elicitation methods, what kind of emotions one wishes to detect, what indicators and features are used for prediction, and the selection of classification algorithms. An increased interest in examining physiological signals has emerged in recent years and is a relatively unexplored field compared to emotion recognition based on facial expression, speech, body gestures, and poses. Some of the most successful studies are based on Electroencephalography (EEG) signals. Still, research shows that peripheral physiological signals such as Electrocardiography (ECG), respiration, Electrodermal Activity (EDA), Electromyography (EMG), and skin temperature can also be valuable indicators of emotions. Yoo et al. [55] proposed a neural network-based method detecting four different emotions by using GSR and ECG as indicators. Yuan-Pin Lin and Chen [57] achieved an accuracy of 82.29% detecting four emotions with EEG measurements as an indicator, and Priyanka Das and Tibarewala [36] have applied binary classification to detect happiness/neutral, happiness/sadness, and neutral/sadness with 91.24%-93.32% accuracy. Table 1.1 gives an overview of some selected studies regarding emotions detection using physiological signals.

Table 1.1: Comparison of 11 studies outlining the measured physiological signals, number of participants, triggered emotions (intent), method used to excite emotions, machine learning classification scheme during analysis, and accuracy of correlation.

Author	Signals	No. of subjects	Emotions	Emotion elicitation method	Classification Scheme	Accuracy Rate
Kim and André [27]	EMG, ECG,SC[1], RSP[2]	3	Q1, Q2, Q3, Q4 in valence/arousal plane	Music	pLDA[3]+EMDC[4]	69.70%
Yuan-Pin Lin and Chen [57]	EEG	26	Joy, anger, sadness, pleasure	Music	SVM	82.29%
Yu-Liang Hsu [56]	ECG	61	Joy, Tension, sadness, peacefulness	Music	SFFS-KBCS[5]+GDA[6] +LS-SVM[7]	61.52%
Priyanka Das and Tibarewala [36]	GSR, ECG	4	Happy, neutral, sad	Video clips	SVM	91.24%-93.32% (Binary)
Wanhui Wen and Huang [53]	OXY[8],GSR,ECG	101	Amusement, Anger, Grief, Fear, Baseline plane	Video clips	Random forest classifier	74%
Gaetano Valenza and Barbieri [18]	ECG	30	valence/arousal	Images from IAPS[9]	LDC[10],QDC[11], KNN[12],PNN[13], VDC[14],MLP[15], SVM	79% / 83.55%
Udovičić et al. [50]	GSR, PPG [16]	13	valence/arousal	Pictures from GAPED[17]	SVM, KNN [12]	67% / 70,3%
Giakoumis et al. [20]	GSR,ECG	19	Boredome	Video games	LDA [18]	85.19%
Rakshit et al. [37]	PPG[15]	33	Happy, sad, neutral, null	Video clips (Harry meets sally, The Champ, neutral +-sign on screen)	SVM	83.8%
Setyohadi et al. [43]	GSR	39	Positive, negative, neutral	Audio-visualisation	SVM	75.65%
Yoo et al. [55]	GSR, ECG	6	Q1, Q2, Q3, Q4 in valence/arousal plane	Video clips (chosen by the subjects)	Neural Network	80.2%

[1] SC: Skin Conductance
[2] RSP: Respiration
[3] pLDA: Probibalistic Linear Discriminant Analysis
[4] EMDC: Emotion-specific multilevel dischotomous
[5] SFFS-KBCS: sequential forward floating selection-kernel-based class separability
[6] GDA: Generalized Discriminant Analysis
[7] LS-SVM: Least Squares Support Vector Machine
[8] OXY: blood oxygen saturation
[9] IAPS: International Affective Picture System
[10] LDC:Linear Discriminant Classifier
[11] QDC: Quadratic Discriminant Classifier
[12] KNN: K-Nearest Neighborhood
[13] PNN: Probabilistic Neural Network
[14] VDC: Vector Distance Classifier
[15] MLP: Multilayer Perception
[16] PPG: Photoplethysmogram
[17] GAPED: Geneva Affective Picture Database
[18] LDA: Linear Discriminant Analysis

1.3 Structure of the thesis

In this chapter we have presented a short introduction to emotion detection, a motivation for developing a system that can be used for recognizing specific emotions, some application areas, and existing literature within the field.

Continuing, the thesis is presented as follows:

Chapter 2 Introduces theory as a foundation for the methodology of the study. In this chapter we discuss the complexity of emotions in a psychological, philosophical, biological and neuroscientific point of view. We present generally accepted emotional models and we describe the mathematics behind the computer scientific approaches that are used for data analysis in Chapter 3. For an expert in the field, this chapter may be skipped.

Chapter 3 Focuses on the chosen methodology for the thesis. This chapter is split into two parts. The first part if the chapter covers the data collection method which include the hardware choices, user interface, method for emotional triggering, and experimental protocol. The second part of the chapter presents the data analysis. This section includes signal processing, feature extraction, PCA for classification, and an application of a SVM with a Radial Basis Fuction (RBF) kernel.

Chapter 4 Presents the experimental results.

Chapter 5 Discusses results and methodology in the light of theory and similar experiments in the literature.

Chapter 6 Concludes the thesis in which the most important findings are presented and suggestions are made for future work.

Chapter 2

Background

Many of the choices that have been made in this thesis are rooted in psychological and philosophical methods and theories in addition to neuroscientific theory and data scientific methods. Due to the scope of the project, this chapter presents background information on a wide range of scientific fields in order to justify the full methodology of the thesis. We discuss the difficulties in defining emotions and the issues of analyzing biosignals across individuals. Moreover, we explain the connection between physiological signals, emotions, and the Autonomic Nervous System (ANS). Furthermore, we present certain standards that are commonly used in emotion studies and will be used for this thesis. Finally, we explain some methods for analyzing the collected bio-signals.

2.1 The Complexity of Emotions

Distinct emotions have proven to be challenging to resolve. We will not try to define emotions in this thesis, but it is essential to understand the complexity of the field and the issues of defining it before we investigate physiological signals connected to emotional reactions.

Emotions have intrigued scientists for a long time, and recognized theories have roots from ancient philosophy. Those roots have given the study of emotions a position in the field of philosophy, even though some will argue that its true position should be in the field of psychology or neuroscience. This limbo between the different fields is not unproblematic which complicates a systematic study of the topic. Furthermore, emotion studies have become an arena for armchair speculation. Everyone has experienced their own emotions, and thus is the temptation of unsystematically, without a theoretical background, presenting theories and ideas about the field. As Descartes said in his introduction of the topic: “Everyone has experience of the passions within himself and there is no necessity to borrow one’s observations from elsewhere in order to discover their nature” [28].

Certain aspects make emotions particularly difficult to define. Thus the topic must be studied carefully.

The Subjectivity of Emotions Darwin [14] argued that emotions are crucial for survival, and thus they have distinctive expressions that should be accurately recognized by all humans. This statement led to a belief that at least some emotions are not learned but universally pre-programmed into our brains. Ekman [17] supported Darwin's hypothesis and defined, based on cross-cultural experiments, six universal, basic emotions: anger, disgust, happiness, sadness, and surprise. According to Ekman, all the six basic emotions can be recognized by facial expression across cultures and borders.

Paul Ekman's work and Darwin's hypothesis have later been criticized by other theorists, including Barrett [6] whom rejects Darwin's hypothesis and describes emotions as physical experiences that each of us constructs based on our unique personal history.

Even if there is a set of basic emotions, there is a consensus in which emotions are subjective to some extent. Thus cultural differences need to be considered in emotion studies. To compensate for this subjectivity, most studies include some form of self-assessment. In this way, the subjectivity of emotions is partly taken care of. However, the self-assessments usually consist of words or pictures which can be subjectively interpreted. Words are even translated between countries. Since some countries have a much richer language, with a wide range of expressions for emotions than other countries, it would be impossible to translate an emotional expression in its complete sense from one language to another. Furthermore, Barrett [6] criticizes the belief that one can recognize an emotion based on facial expressions as the connection between an emotion and a facial expression is highly individual. This statement weakens the use of images in the self-assessments. In fact, some argue that the only way to describe an emotion is to compare it with another emotion one had in another situation [28].

Unconscious Emotions There are some unconscious aspects regarding emotions that must be considered. You may have been in a situation where you know that you are feeling something. Still, you cannot to define the emotion. Alternatively, you may have heard stories about people who would never think that they were depressed until they went to a psychologist and were diagnosed. John F. Kihlstrom et al. [26] calls this emotional effect "implicit emotions" and describes them as the effect of not recognizing/experiencing an emotional state even though an emotional response, such as an automatic facial expression or an increased heart rate, is generated.

Emotion Regulation Philosophers have discussed emotions and reason as two conflicting elements. Aristotle believed that emotion had an inferior role, that it was less intelligent, more primitive, and dangerous than reason, and thus it needed to be controlled by reason. While David Hume (1739/1888) declared that "Reason is, and ought to be, the slave of the passions." Today, there is a greater consensus among psychologists and philosophers that reason and emotion are highly connected and that the brain is just predicting how to react based on past experiences and similar situations. Despite having this knowledge, we continually regulate our emotions by hiding a smile when winning a card game not to hurt our friends, or by suppressing negative emotions to avoid the pain ourselves. A dramatic increase in research regarding emotion regulation has occurred in the last 40 years. One sees the great advantage of taking control of one's own emotions both in a social setting and for one's own happiness.

Emotion regulation may complicate emotion studies in which the participants know that they are in a test situation and can prepare for control. Furthermore, there is a large gap in people's ability to regulate emotions, making it difficult to compare the participants' reactions.

2.1.1 Valence-Arousal Plane

Even though one can discuss to what extent discrete emotional labels like words are useful in self-assessments, one can agree that it will capture some variance. Furthermore, there will always be a need for a model to study the field. Russell [38] suggested that rather than expressing affect by monopolar factors, affect is better described by a few independent bi-polar dimensions. Russell proposed a two-dimensional, Circumplex model (see figure 2.1) which is a plane spanned by the two axis: valence; whether an emotion is positively or negatively loaded; and arousal; the intensity of the emotion. Thus the valence-arousal plane provides a continuous scale which captures a much greater range of emotional states than a discrete notion could ever manage.

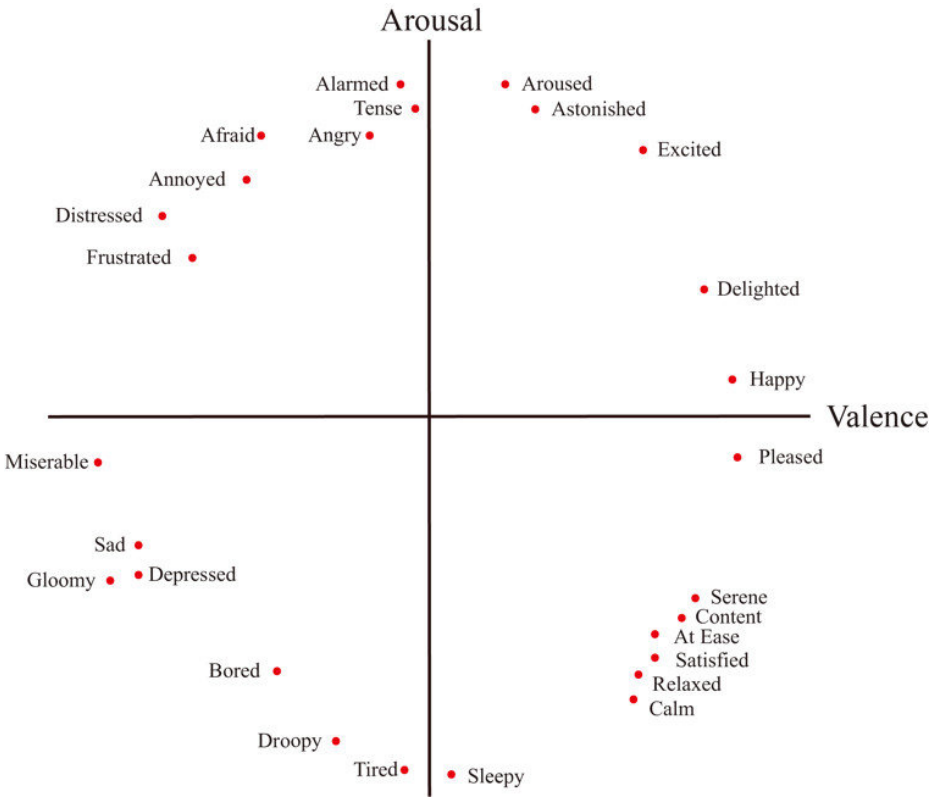


Figure 2.1: The Circumplex model developed by Russell [38] including the postulated basic emotions. The model is spanned by an arousal- and a valence axis and the idea is that every emotional state can be placed in this plane. Image courtesy of [42].

2.1.2 Self Assessment Manikin

Inspired by Russell [38], Bradley and Lang [9] developed a picture-oriented questionnaire that was easy to use, and that was independent of language. In addition to valence and arousal, the SAM added control as a third dimension to describe emotions. We are, however, in this thesis focusing on the two dimensions valence and arousal for the reason that we can combine the Bradley and Lang [9]'s SAM and Russell [38]'s circumplex model. Thus, we use the two panels depicted in figure 2.2 for self-assessment.

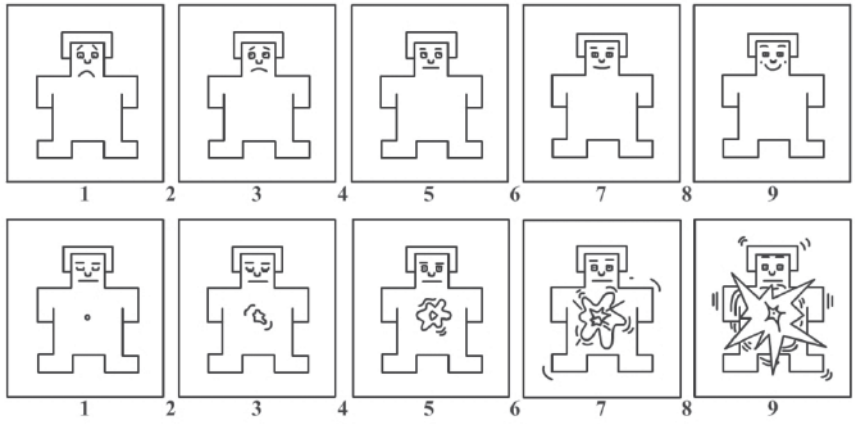


Figure 2.2: The SAM developed by Bradley and Lang [9] for a common picture-based questionnaire to include in emotional studies. The first row is related to the valence axis indicating a positive or negative feeling. The second row indicates arousal; the larger the star, the more intense the emotion and vice-versa. Notice that SAM originally contains the three panels: valence, arousal and control. In this experiment we neglect the control panel. Image courtesy of [45].

2.2 Autonomic Nervous System

The Autonomic Nervous System (ANS) is a control system in our body that acts largely unconsciously and regulates heart rate, respiratory rate, and certain hormone production, among other bodily functions. The autonomic nervous system has proven to be the main component in controlling the fight-or-flight response, which is a commonly associated

with the emotion *fear*. The body is getting ready for fight or flight by moving tension and energy from organs that are not important in a threatening situation to the organs that are important. Besides the emotion *fear*, there is no consensus that the ANS responds to all other types of emotions. However, studies have generated good results regarding the topic as mentioned in section 1.2 page 2. ANS has two branches called the Sympathetic Nervous system (SNS) and the Parasympathetic Nervous System (PSNS) which have opposite functions in the body. The SNS prepares the body for a fight, while the PSNS acts like a break.

2.2.1 Sympathetic Nervous System

The SNS is considered the fight-or-flight system. Some typical sympathetic nerve activity includes increased respiratory- and heart rate, decreased digestive system activity, increased blood glucose level, increased adrenaline, and perspiration.

2.2.2 Parasympathetic Nervous System

The PSNS is known as the rest-and-digest system. The PSNS acts in opposition to the SNS by decreasing the heart- and respiratory rate, and increasing activity in the digestive system.

2.3 Measurable Physiological Signals

When designing a system that dynamically integrates emotional states, it is important to first understand the measuring physiology. Physiological sensors have been developed based on measuring the responses for the sympathetic and parasympathetic nervous systems. In an engineering context, there exist many sensors for measuring physiological signals such as Electrocardiography, Galvanic Skin Response, Electromyography, Electroencephalography, Body Temperature, Pulse Oximetry, and Blood Pressure. We expand on only the measurement tools used in the project solution.

2.3.1 Galvanic Skin Response

The Galvanic Skin Response (GSR) falls under the umbrella of Electrodermal Activity. It is used to detect autonomic changes in the skin's electrical properties by measuring the skin conductance.

The GSR sensor is commonly used in lie detectors and studies where emotional behavior is analyzed. The use of GSR sensors for these purposes is based on the following two assumptions:

1. The conductivity of the skin depends on the state of sweat glands.
2. The sympathetic nervous system regulates the sweat gland.

Several studies have shown that Galvanic Skin Response is a good indicator of emotional arousal. However, it is still uncertain whether GSR reveals information about emotional

valence. Nevertheless, the GSR is a significant measure in which it is the only autonomic psycho-physiological variable that is not contaminated by parasympathetic activity [16].

2.3.2 Electrocardiography

The heart is one of the organs affected by the automatic nervous system, and thus heart activity has been extensively researched in connection with affection studies. An Electrocardiography (ECG) is widely used in heart analysis and has been an important tool for detecting heart properties and defects.

An ECG is a graph of the electrical activity of the heart. It is produced by placing electrodes at the skin detecting small electrical changes caused by a depolarization followed by repolarization of the cardiac muscle during each cardiac cycle. There are three main ECG features that indicate the different phases of cardiac electrical activity, these are

P wave atrial depolarization

QRS complex ventricular depolarization

T wave ventricular repolarization

Figure 2.3 shows a standard ECG waveform.

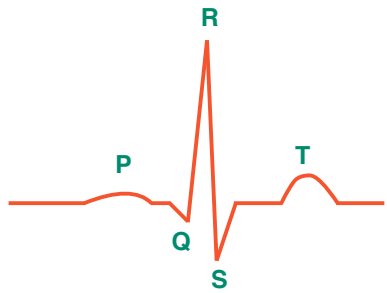


Figure 2.3: The ECG waveform: the P wave indicates atrial depolarization, the QRS complex shows the ventricular depolarisation, and the T wave occurs during ventricular repolarization.

Heart Rate Variability

By measuring the time between each R peak of the ECG signal (see Figure 2.3), one obtains the RR-intervals, which indicate the Heart Rate Variability (HRV). Heart rate is affected by both the SNS and the PSNS caused by the continuous change in the sympathetic-parasympathetic balance, thus causing fluctuations to occur around the mean heart rate. It has been shown that the fluctuations are related to the respiratory rate, the baroreflex, and thermoregulation [52]. The HRV has frequently been used as a measure for stress to determine whether the PSNS or the SNS is dominant at a certain time. However, it is important to note that there are large differences between the HRV of individuals. For example, one

has seen a much stronger connection between the respiratory rate and the HRV in young athletes than older and less healthy persons. An RR-interval is calculated as shown in Figure 2.4.

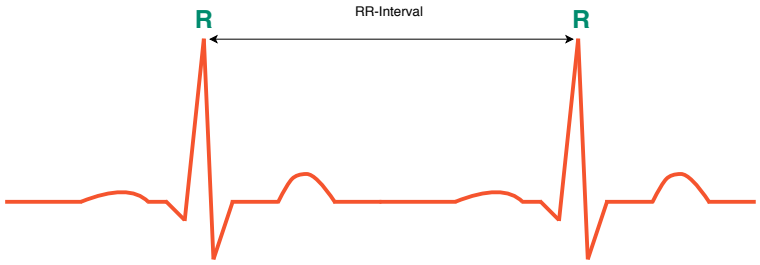


Figure 2.4: RR-intervals measure the time interval between peaks to indicate HRV.

Time Domain Analysis HRV analyzed in the time domain can be divided into the two indices short-term variability and long-term variability. Short-term variability indices represent a fast change in heart rate and are related to the respiratory system, whereas long-term variability indices are slower fluctuations that are caused by the baroreflex and thermoregulation [52].

Frequency Domain Analysis Spectral analysis as a method for studying HRV was introduced by SAYKRS [39], and has since then been included in a majority of research containing HRV analysis. Spectral analysis enables a study of frequency-specific oscillations, which is not accessible in the time domain. There are methods for filtering out sympathetic- and parasympathetic activity by dividing the power spectrum of recorded HRV into the three frequency bands: very-low-frequencies, low frequencies, and high frequencies. Furthermore, it has been shown that the low-frequency band and the high-frequency band are related to the SNS and the PSNS, respectively. One has even seen that the ratio between high- and low- frequency heart rate variability is stable with advancing age [41]. This is a very favorable property when studying the balance between the sympathetic- and parasympathetic nervous system across individuals.

2.3.3 Body Temperature

Normal body temperature is typically in the range of 36.5°C-37.5°C, and the body has an advanced regulating system to keep the body temperature within that range, regardless of the ambient temperature. Nevertheless, there are small variations in the body temperature that depend on different factors, including one's emotional state.

2.4 The Complexity of Biosignals

We have already discussed why emotions are difficult to define from a philosophical and psychological perspective. From a biological, neuroscientific point of view, the topic be-

comes more complex. Biosignals are commonly known to have a low signal-to-noise ratio affected by internal and external artifacts. Furthermore, a persons' physiological responses are influenced by many factors.

First of all, the sensors will have small differences in the placement on the body. Body temperature may be affected by ambient temperature, and the GSR can be affected by humidity.

The internal factors are many. We have already mentioned that age and activity level affect Heart Rate Variability. Moreover, inheritance is an important factor for both heart activity and sweat production. Food, drinks, medications, diseases, sex, time of the day and the menstrual cycle are other factors that affect body temperature, heart activity and sweat production.

Some researchers have compared the performance of multi-user classification and single-user classification based on bio-signals. Udovičić et al. [50] shows an accuracy of 86.7% with a single-user model, compared to a multi-user model with 66.7% accuracy. This gap is possibly, partly caused by the issues described in section 2.1, but there is no doubt that biosignals are extremely difficult to analyse across individuals and that such classification is challenging, some will argue, even impossible.

2.5 Signal Processing

Signal processing is commonly a very important step in a machine learning project, in which it may produce more accurate classification. A preprocessing step includes methods for removing noise, scaling methods to make the features comparable, and data transformation methods in order to access certain features of interest. This section presents the preprocessing methods that we will use in this thesis.

2.5.1 Moving-median Filtering

Moving-median filtering is a nonlinear signal processing technique useful for noise suppression. Moving-median filtering is performed by a sliding window replacing the window's middle value with the window's median value. Some advantages of applying a median filter are that it preserves sharp edges and is very efficient for smoothing spiky noise. Figure 2.5 shows an example of how a moving-median filter works on an observed set of data.

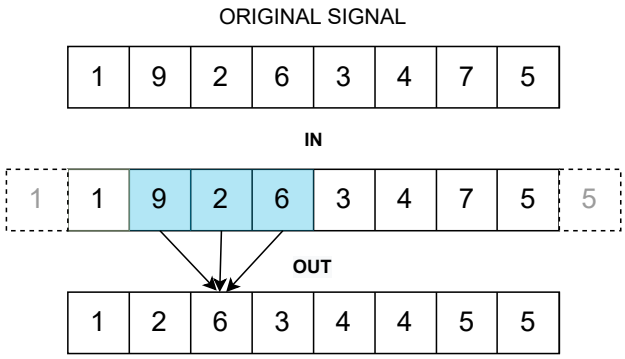


Figure 2.5: An interpretation of moving-median filter with a window size of 3. The middle value of the window is replaced by the window's median value.

2.5.2 Feature Scaling Methods

The range of raw data may vary to a large extent. Since many classifiers calculate Euclidean distances between data samples, normalization methods are needed to create a common scale for the data set.

Min-max Scaling

Min-max scaling is a simple normalization method that scales all values to the range $[0, 1]$ by the following formula

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}\,, \tag{2.1}$$

where x is the observed data.

Standardization

Features often contain measurements with different units. Many machine learning models require a standardization of the data set by rescaling the data to have zero mean and unit variance. The following formula is used for calculating the standardized values

$$x' = \frac{x - \bar{x}}{\sigma}\,, \tag{2.2}$$

where x is the observed data, $\bar{x}$ is the sample mean, and σ is the sample standard deviation.

2.5.3 Power Spectral Density and Welch's Method for Power Spectral Density Estimation

Spectral analysis considers the problem of determining the power over frequency of a time series. Tools from spectral analysis will be used in this thesis to compute opportune features from raw signals. The features are then used as inputs to estimate specific emotions of the participants in the experiment.

From a mathematical perspective, if a time series is finite, one can compute the Power Spectral Density (PSD) of that time series as

$$S_{xx}(f) = \lim_{T\to\infty} \frac{1}{T}|\hat{x}_T(f)|^2 \tag{2.3}$$

$\hat{x}_T(f)$ is the Fourier transform of the time signal $x_T = x(t)w_T(t)$ in which w_T is unity within the interval (-T,T) and zero elsewhere.

A common way to calculate the PSD is to apply the Fast Fourier Transform (FFT). However, in this thesis we rather estimate a PSD by Welch's method to obtain a smoother, stationary signal.

The idea behind applying Welch's method for PSD estimation is to divide the time signal into successive blocks of a chosen length R. A periodogram is calculated for each of the blocks and then averaged for a smoother signal.

Let the mth zero-padded segment from a signal x be denoted by

$$x_m(n) \triangleq w(n)x(n+mR), \quad n = 0, 1, ..., M-1, \qquad m = 0, 1, ..., K-1\,, \tag{2.4}$$

in which window function $w(n)$ contains M nonzero samples and K is the number of available segments. The periodogram of the mth block is given by

$$P_{x_m,M}(\omega_k) = \frac{1}{M}|\mathrm{FFT}_{N,k}(x_m)|^2 \triangleq \frac{1}{M}\left|\sum_{n=0}^{N-1} x_m(n)e^{\frac{-j2\pi nk}{N}}\right|^2 \tag{2.5}$$

and Welch's estimate for Power Spectral Density (PSD) is formulated by the following equation

$$\hat{S}_x^W(\omega) \triangleq \frac{1}{K}\sum_{m=0}^{K-1} P_{x_m,M}(\omega_k). \tag{2.6}$$

There are a number of different window functions to choose from. We are in this thesis using the Hann function defined as following

$$w(n) = \begin{cases} 0.5 + 0.5\cos\left(\frac{2\pi n}{N}\right), & |n| < \frac{N-1}{2}, \\ 0, & \text{otherwise.} \end{cases} \tag{2.7}$$

In this thesis we will especially make use of (2.6) to generate features from raw data. For more information about this type of estimators, we send the interested reader back to Solomon [46].

2.6 Principal Component Analysis

PCA is used in a broad spectrum of applications, but it is especially well utilized within multivariate data analysis. It has many areas of application including visualization, regression, image analysis, data compression, pattern recognition, and time series prediction [49].

Multivariate data analysis can be cumbersome as one must often analyze data with high dimensions. Once the data reaches a dimension higher than three, there is no simple, intuitive way to visualize the relationships between the variables directly.

The idea behind PCA is to decide the most important information in the data set and then reduce the dimension while preserving as much of the data's variation as possible. Thus, it gives the possibility to visualize and analyze a representative and simplified version of the data set.

To give an intuition behind the principle of Principal Component Analysis (PCA), a geometrical interpretation of PCA is presented in figure 2.6. The figure shows an example with only two classes and two features for a simple visualisation of the steps. Also, note that the data is standardized by the method described in section 2.5.2. This is often an important step for PCA. The features may have different units and thus, one must adopt standardization techniques to avoid distortions in the variances.

2.6.1 Principal Component Analysis and Singular Value Decomposition

From its invention in 1901, various methods for computing the principal components have been found. This includes sparse PCA, nonlinear PCA, and robust PCA. However, the most common, standard interpretation of PCA was formulated in 1933 by Harold Hotelling which we present here.

For a set of observed d-dimensional data vectors

$$\mathbf{t}_n \in \mathbb{R}^d \text{ for } n \in 1, ..., N \tag{2.8}$$

with the sample covariance matrix

$$\mathbf{S} = \frac{1}{N-1} \sum_{n=1}^{N} (\mathbf{t}_n - \bar{\mathbf{t}})(\mathbf{t}_n - \bar{\mathbf{t}})' = \frac{1}{N-1} \mathbf{T}'\mathbf{T}\,, \tag{2.9}$$

where

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}_1' - \bar{\mathbf{t}}' \\ \mathbf{t}_2' - \bar{\mathbf{t}}' \\ \vdots \\ \mathbf{t}_N' - \bar{\mathbf{t}}' \end{bmatrix} \tag{2.10}$$

and $\bar{\mathbf{t}}'$ is the transposed data sample mean.
The direction the data set is being projected into is determined by the principal axes

$$\mathbf{w}_j, j \in \{1, ..., q\}\,. \tag{2.11}$$

Deciding the first principal component is done by maximizing $\mathbf{w}_1^T \mathbf{S} \mathbf{w}_1$.
This maximizing problem can be solved by the method of Lagrange multipliers letting

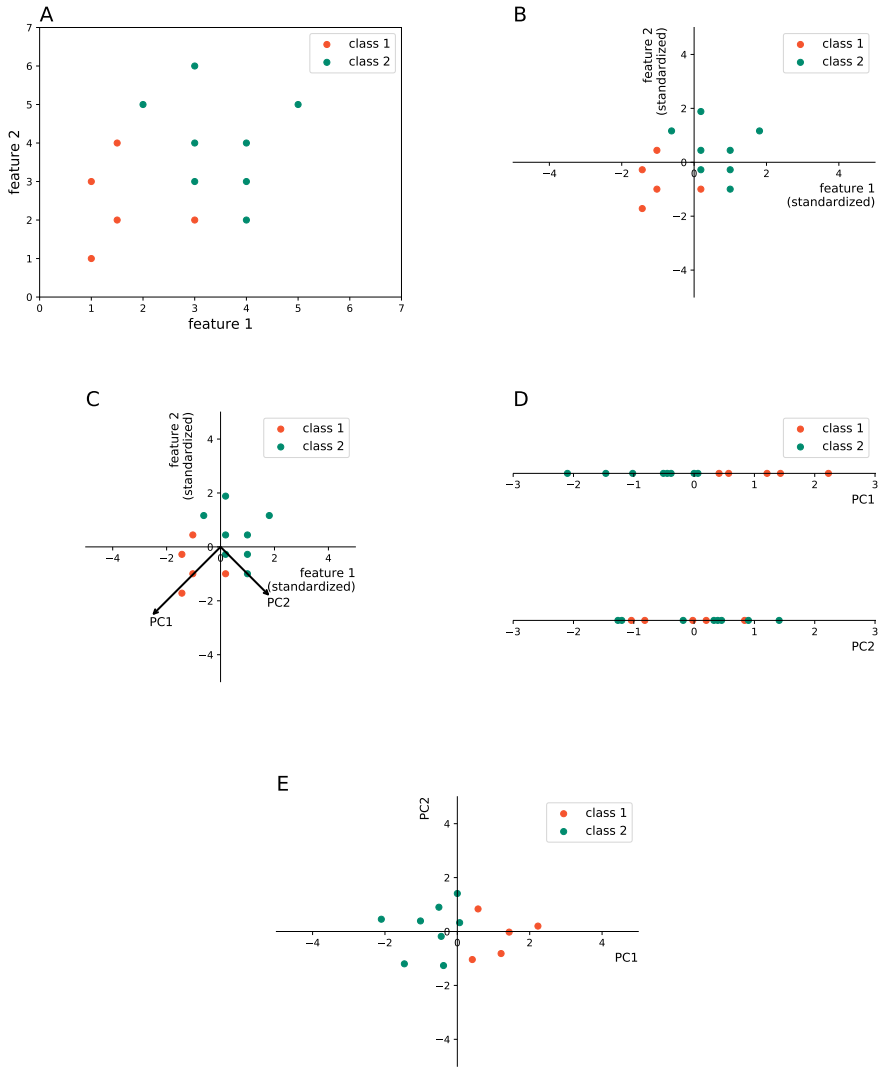


Figure 2.6: (A) Each dot represents a data point in its raw form, expressed by two features. The data belongs to two different classes and it is colored accordingly. (B) The data is standardized by removing the mean and scale to unit variance. (C) PCA identifies the two directions(PC1 and PC2) along which the data has the largest variance. (D) The data is projected onto each of the Principal Components(PC1 and PC2). (E) Two principal components now express the dataset by rotating the axis and project the data points from the 1-dimensional axis PC1 and PC2 onto a plane spanned by the two Principal Components.

$$\mathbf{S}\mathbf{w}_1 = \lambda_1 \mathbf{w}_1 \tag{2.12}$$

The next principle components are calculated by the same procedure with the additional constraint that the principal component that is being calculated has to be perpendicular to the previous calculated principal components.

Singular Value Decomposition

Singular Value Decomposition (SVD) is a matrix decomposition method with countless applications. However, for Hotelling's PCA, SVD plays a very central role as it constitutes the main principle behind the calculations of the principal components.

Reduced form SVD of a general $m \times n$ matrix $\mathbf{A}$ can be written as

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}' = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i' \tag{2.13}$$

where

$$\mathbf{U} = \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & ... & \mathbf{u}_r \\ | & | & & | \end{bmatrix}, \mathbf{V} = \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & ... & \mathbf{v}_r \\ | & | & & | \end{bmatrix} \tag{2.14}$$

and have orthogonal columns,

$$\mathbf{U}'\mathbf{U} = \mathbf{I}_{(r)} \tag{2.15}$$

$$\mathbf{V}'\mathbf{V} = \mathbf{I}_{(r)} \tag{2.16}$$

$r = (\mathbf{A})$. $\mathbf{\Sigma}$ is a $r \times r$ diagonal matrix that contains the singular values, σ_i, of $\mathbf{A}$. The singular value is the square root of the non-negative eigenvalues of $\mathbf{A}'\mathbf{A}$ or $\mathbf{A}\mathbf{A}'$.

Let $\mathbf{T}$ from equation (2.9) be decomposed by SVD.
Simple math gives

$$\mathbf{T}'\mathbf{T} = (\mathbf{U}\mathbf{\Sigma}\mathbf{V}')'\mathbf{U}\mathbf{\Sigma}\mathbf{V}' = \mathbf{V}\mathbf{\Sigma}'\mathbf{U}'\mathbf{U}\mathbf{\Sigma}\mathbf{V}' = \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}' \tag{2.17}$$

From equation (2.9) we have that

$$(N-1)\mathbf{S} = \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}' \tag{2.18}$$

One can thus observe the following relationship between the eigenvalues of the principal components in equation (2.12) and the singular values of the SVD by

$$\lambda = \frac{\sigma^2}{N-1}, \tag{2.19}$$

Applying SVD to compute the principle components is an efficient approach in which it can handle sparse matrices and it even exists reduced forms of SVD which are economic to compute.

Scores and loadings

We are in this thesis vizualizing our data set by following

1. projecting the data samples onto the low-dimensional sub-space, spanned by principal components, presented in a score plot. The scores is given by

$$\mathbf{X} = \mathbf{TW} \tag{2.20}$$

   where $\mathbf{W}$ is a matrix with columns that are eigenvectors of $\mathbf{T}'\mathbf{T}$.

2. calculating the loadings i.e. which features contributes to the different principal components. This is simply done by investigating the cross-covariance between the original features and the standarized principal components. For this thesis we are using heatmaps to present the loadings.

Principal Component Analysis is sometimes used in combination with classification algorithm in which a chosen number of principal components are used instead of the original data vectors. In this thesis we combine PCA with Support Vector Machines (SVM).

The next section describes the ideas behind Support Vector Machines with mathematical derivations, and some extensions to the original algorithm which we will use in this thesis.

2.7 Classification by Support Vector Machines

Support Vector Machines (SVM) are supervised learning models for classification, regression, and outlier detection.

The main idea behind classification with SVM is to find an optimal separating hyperplane based on a set of training data, which can later be used to classify new, unknown test data. SVM is generally defined as a non-probabilistic binary linear classifier as it, based on the hyperplane placement, assigns new samples to one class or the other. Although we will later see that we can expand this definition by applying different methods.

SVM differ from other classification algorithms by the way they decides the hyperplane based on the most similar data points across the different classes. Those data points are called support vectors. Figure 2.7 shows an example of the optimal hyperplane that separates two different classes expressed by two features. For this special case, the hyperplane is reduced to a simple line. The hyperplane is placed such that the margin between the hyperplane and the support vectors is maximized.

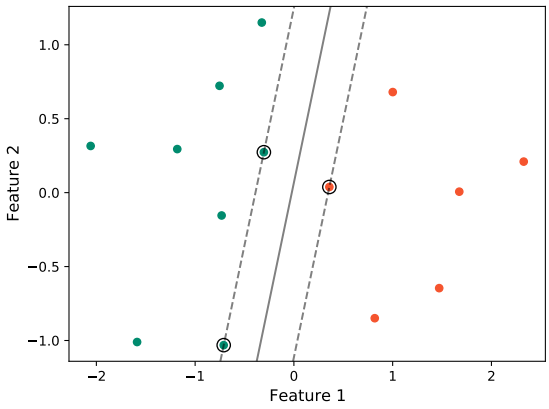


Figure 2.7: Each dot represents a data point expressed by two features. The dataset consists of two classes, and the dots are colored according to which class it belongs to. The support vectors are highlighted with a black circle. The support vectors decide the optimal hyperplane represented as a gray solid line. The dotted line is the margin, which is maximized by placing the hyperplane in such a manner.

2.7.1 A Mathematical Formulation

Let $\mathbf{x}_i \in \mathbb{R}$ for $i = 1, .., n$ be training vectors that represent two classes such that the training samples are linearly separable and let $\mathbf{y} \in \{1, -1\}^n$ be a vector with the corresponding labels.

The general formula for a hyperplane is through the point x formulated as follows

$$\mathbf{w}'\mathbf{x} + b = 0\,, \tag{2.21}$$

where $\mathbf{w}$ is a normal vector to the hyperplane.

Let $\mathbf{x}$ be an arbitrary training sample, let $\mathbf{x}^p$ be the projected training sample onto the hyperplane and let $\mathbf{d}$ be the distance between $\mathbf{x}$ and $\mathbf{x}^p$ as shown in figure 2.8.

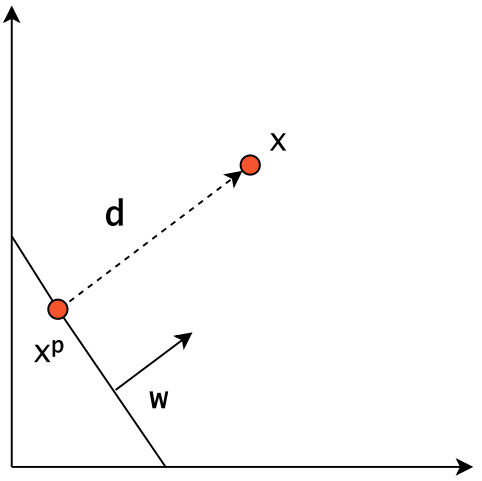


Figure 2.8: Hyperplane with defined variables. $\mathbf{x}$ is an arbitrary training sample. $\mathbf{x}^p$ is the projected training sample onto the hyperplane. $\mathbf{d}$ is the distance between the training sample and the projected training sample, and $\mathbf{w}$ is a vector normal to the hyperplane.

This hyperplane, through the projected point $\mathbf{x}^p$, can then be expressed as

$$\mathbf{w}'(\mathbf{x} - \mathbf{d}) + b = 0 \tag{2.22}$$

where $\mathbf{x} - \mathbf{d} = \mathbf{x}^p$. Since $\mathbf{w}$ and $\mathbf{d}$ are parallel, $\mathbf{d}$ can be expressed as

$$\mathbf{d} = \alpha\mathbf{w} \tag{2.23}$$

in which α is a constant. Replacing $\mathbf{d}$ in equation (2.22) gives

$$\mathbf{w}'(\mathbf{x} - \alpha\mathbf{w}) + b = 0 \tag{2.24}$$

Solving for α gives

$$\alpha = \frac{\mathbf{w}'\mathbf{x} + b}{\mathbf{w}'\mathbf{w}}\,. \tag{2.25}$$

By replacing α in equation (2.23), we get the following expression for the distance vector $\mathbf{d}$

$$\mathbf{d} = \frac{\mathbf{w}'\mathbf{x} + b}{\mathbf{w}'\mathbf{w}}\mathbf{w}\,, \tag{2.26}$$

and thus the length of $\mathbf{d}$ is

$$||\mathbf{d}|| = \sqrt{\mathbf{d}'\mathbf{d}} = \frac{|\mathbf{w}'\mathbf{x} + b|}{||\mathbf{w}||}\,. \tag{2.27}$$

The margin γ can be found by minimizing the length of $\mathbf{d}$ such that

$$\gamma(\mathbf{w}, b) = \min_{\mathbf{x}} \frac{|\mathbf{w}'\mathbf{x} + b|}{||\mathbf{w}||}\,. \tag{2.28}$$

To ensure that all training samples lie on the correct side of the hyperplane, the following constraint is added to the minimization problem:

$$y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 0 \quad \forall i \tag{2.29}$$

SVM require the margin to be maximized. However, maximizing a minimization problem can be cumbersome. Thus, letting

$$|\mathbf{w}'\mathbf{x} + b| = 1 \tag{2.30}$$

which gives

$$\max_{\mathbf{w},b} \frac{1}{||\mathbf{w}||} \cdot 1 = \min_{\mathbf{w},b} ||\mathbf{w}|| = \min \mathbf{w}'\mathbf{w} \tag{2.31}$$

and we can add another constraint to our problem to compensate for equation (2.30) where we can then write the complete optimization problem as

$$\begin{aligned} \min_{\mathbf{w},b} \quad & \mathbf{w}'\mathbf{w} && \text{(2.32a)} \\ \text{s.t.} \quad & \forall i \quad y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 0, && \text{(2.32b)} \\ & |\mathbf{w}'\mathbf{x} + b| = 1\,. && \text{(2.32c)} \end{aligned}$$

By studying the two constraints, one can observe that they can be written as one simple constraint. The final optimization problem is then

$$\begin{aligned} \min_{\mathbf{w},b} \quad & \mathbf{w}'\mathbf{w} && \text{(2.33a)} \\ \text{s.t.} \quad & \forall i \quad y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1\,. && \text{(2.33b)} \end{aligned}$$

2.7.2 Support Vector Machines with Soft Margins

There are situations where a hyperplane separating the classes does not exist. In those cases in which the dataset is non-separable, there will not be a solution to the optimization problem (2.33). Cortes and Vladimir [13] introduced SVM with soft margins. The idea was to introduce slack variables allowing the constraints to be slightly violated and thus a solution exists even for non-separable training data. The optimization problem with slack variable ξ_i is presented by equation (2.34). The parameter C controls a tradeoff between sacrificing some points by using a large margin and letting the margin be small such that the data points are much closer to stay on the correct side of the hyperplane. The optimisation probelem then becomes

$$\begin{aligned} \min_{\mathbf{w},b} \quad & \mathbf{w}'\mathbf{w} + C\sum_{i=1}^{n} \xi_i && \text{(2.34a)} \\ \text{s.t.} \quad & \forall i \quad y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1 - \xi_i, && \text{(2.34b)} \\ & \forall i \quad \xi_i \geq 0\,. && \text{(2.34c)} \end{aligned}$$

A closed-form expression for ξ_i can be written as

$$\xi_i = \max[1 - y_i(\mathbf{w}'\mathbf{x}_i + b), 0]\,. \tag{2.35}$$

Equation (2.34) can thus be written as

$$\min_{\mathbf{w},b} \quad \mathbf{w}'\mathbf{w} + C\sum_{i=1}^{n} \max[1 - y_i(\mathbf{w}'\mathbf{x}_i + b), 0] \tag{2.36}$$

which is unconstrained and consists of a regularizer and a hinge-loss.

2.7.3 Kernelized Support Vector Machines

The formulations above consider linear classification problems. This causes a strong bias problem in which they are only able to train linear separable data. Boser et al. [8] reduced this bias for SVM by introducing kernels. The idea is to map the original feature space into a high-dimensional feature space.

$$\mathbf{x} \rightarrow \phi(\mathbf{x}) \tag{2.37}$$

However, this mapping naturally comes with an extremely high computational cost. The kernel trick is thus introduced to avoid this high cost. The idea behind the kernel trick is to never represent the data directly by this mapping but instead to preserve relationships in the high-dimensional by computing a kernel function, which is the inner product of the mapped data.

$$K(\mathbf{x}, \mathbf{z}) = \phi(\mathbf{x}')\phi(\mathbf{z}) \tag{2.38}$$

The Radial Basis Function Kernel

Many different kernel functions have been derived from equation (2.38) by the rule that the function K must be positive semi-definite. One of the most popular kernels is the Radial Basis Fuction (RBF) kernel, defined by

$$K(\mathbf{x}, \mathbf{z}) = e^{-\frac{(\mathbf{x}-\mathbf{z})^2}{\sigma^2}}\ . \tag{2.39}$$

The RBF-kernel maps the feature space into an infinite feature space. The kernel function behaves like a weighted nearest neighbour model in which, in practice, classifies new observations based on how the closest data points are labeled.

2.7.4 Multi-class Classification with Support Vector Machines

As mentioned in the beginning of this section, SVM are, in general, binary classification algorithms. However, there are several techniques for solving multi-class problems with SVM. One popular method is to reduce the problem to multiple binary classification problems and then apply a one-versus-one approach, in which $\frac{N_{\text{classes}} \times (N_{\text{classes}}-1)}{2}$ classifiers are constructed and trained on data from combinations of two classes. The test data is then classified based on "votes" from the classifiers.

Chapter 3

Methodology

The chosen method for this thesis is comprehensive and can be divided into many smaller phases. In this chapter, we will present the method in two parts. The first section covers the data collection which includes all of the hardware and software choices for building the sensing system, selection of elicitation method and the protocol for collecting the bio-data. The second part of this chapter considers the process of analyzing the collected data in which chosen signal-processing methods, feature extraction and classification models are presented.

3.1 Data Collection

The majority of the choices that we take in this study concern the methods for data collection. This includes selecting a sensing systems, a well-developed user interface, an elicitation method for triggering different emotions, a limited number of specific emotions we endeavour to trigger, and an experimental protocol. This section will guide you through these choices and provide a thorough description of the background for our choices.

3.1.1 Hardware Choices

There are numerous biosensors on the market, and there is a great variation in what sensors have been used in previous, similar experiments. Many of the sensors that has been used are massive and expensive [27, 56, 18, 20, 55]. Udovičić et al. [50] have based their experiment on cheaper, wearable equipment by using a compact Shimmer3 sensor. However, Udovičić et al. [50] do not seem to reach the accuracy obtained using more expensive equipment. We want to propose an alternative set of sensors; as compact as the Shimmer3 sensor; with an even lower cost.

Pulse Sensor: Polar H10

Today, there are many low-cost pulse sensor devices for training and workouts. A great motivation for choosing the Polar H10 with a Polar Pro strap for this experiment is that it could communicate with an application directly. The communication of many pulse belts from other manufacturers goes via a smartwatch during or after the workout. Choosing such a pulse belt would firstly increase the price largely in which a watch is much more expensive than a pulse belt, and secondly, the data transmission would be more complex.

Polar H10 is a precise pulse sensor that transmits data via Bluetooth® and ANT+™ (wireless technology that allows monitoring devices to communicate). The pulse data is detected by two electrodes attached to the skin. The electrodes detect ECG signals from the skin, which is first filtered to reject noise and disturbances and then processed to detect the QRS complex. The Polar H10 is thus able to calculate the RR-intervals and the heart rate (Beats per minute (BPM)) in addition to collecting the raw ECG signal.

Oy [33] have conducted an analysis of the accuracy of different pulse sensors, in which the Polar H10 belt is said to have an accuracy of 99.3% when detecting RR-intervals. In fact, according to Oy [33], Polar H10 shows better accuracy than three different Holter monitors (portable medical device for cardiac monitoring).

Figure 3.1: Polar H10 pulse sensor with Polar Pro strap used in the experiments. Image from [35].

GSR Sensor: Grove-GSR Sensor

The market for commercial, low-cost GSR sensors is small. For EDA measurements we choose an arduino-based GSR sensor from Grove. To use an arduino as a platform in the project gives the advantage of compactness in which multiple sensors can be connected to the same arduino. Moreover, the arduino can easily communicate through Bluetooth® by including a Bluetooth® module.

The Grove GSR sensor comes with a pair of finger straps for placing two Ni electrodes on two different fingers. The sensor is based on a constant voltage system in which the GSR sensor applies a constant voltage to the electrodes. The resistance can be found by Ohms law measuring the current flow from one electrode to the other. The human skin resistance can be calculated from the port readings by the following equation

$$\text{Human skin resistance} = \frac{(1024 + 2 * \text{Serial\_Port\_Reading}) \times 10000}{(512 - \text{Serial\_Port\_Reading})} \tag{3.1}$$

Figure 3.2 shows the electronic schemiatic and figure 3.3 shows the full set of components of the Grove GSR sensor.

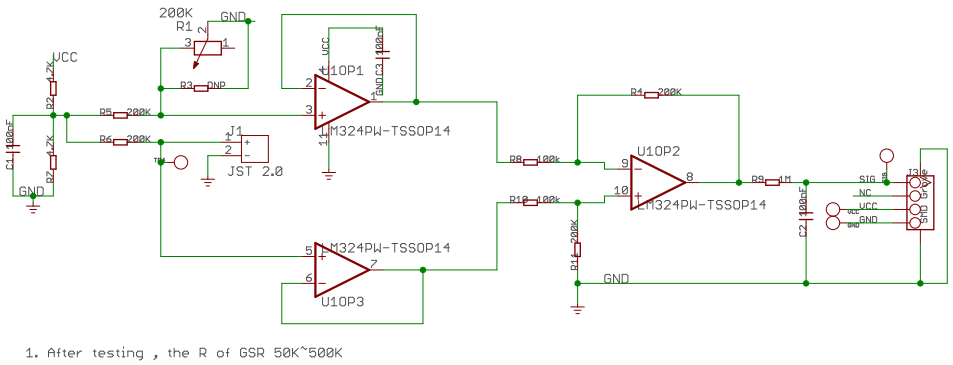


Figure 3.2: Sechematics of the Grove GSR sensor. Image courtesy of [47].

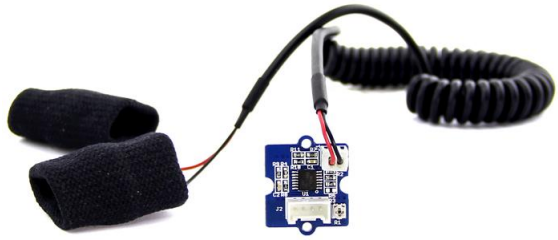


Figure 3.3: Grove GSR sensor used in the experiments. Image courtesy by [47]

Digital Thermometer: DS18B20

Measuring body temperature is challenging since ambient temperatures often cause miscalculations. The most accurate method for measuring the body temperature is to measure the oral, rectal, or ear temperature. However, these measurement methods are invasive and require a different sensor for each participant which is expensive and cumbersome.

Thus, instead one may place a small integrated circuit in the armpit; inspired by Md.Yassin and Chin [30]. Such a small sensor can be easily attached to a strap for the arm so that the sensor is well placed under the armpit. For this purpose, we use the digital thermometer DS18B20.

The DS18B20 is a digital thermometer that provides 9-bit to 12-bit Celsius measurements. The thermometer communicates over a 1-wire bus, which means that it requires only one data line for communication with a central microprocessor. In parasite power mode, the sensor can operate with only one data pin in addition to ground. We are however, for this experiment, using an external power supply by including a V_{DD} pin.

A 1-wire network consists of a single open drain wire that requires a pull-up resistor for communication. The desirable resistance for the DS18B20 is 4.7 KΩ.

The DS18B20 can measure temperatures from $-55°$C to $+125°$C and has an accuracy of $\pm 0.5°$C in the temperature range $-10°$C to $+85°$C.

A schematic of the DS18B20 is in figure 3.4 and an image of the digital thermometer is shown in figure 3.5.

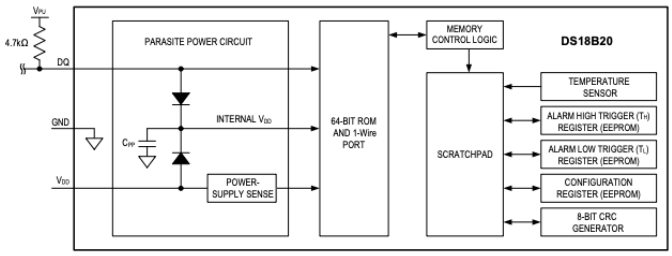


Figure 3.4: The figure shows a block diagram of the DS18B20. A strechpad memory contains the temperature register that stores the digital output of the sensor. A unique serial code of the device is stored in the 64-BIT ROM. The thermometer also contains registers for user-defined alarm trigger values T_H and T_L. Image courtesy by [25]

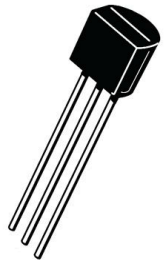


Figure 3.5: Digital thermometer DS18B20 used for the expiriment. Image courtesy by [15]

3.1.2 Seeeduino Setup

Seeeduino V4.2 is an Arduino compatible board with a ATmga328P MCU, an Arduino UNO bootloader, and an ATMEGA16U2 UART-to-USB converter. The Seeeduino board is developed by Seeedstudio.

The setup for the GSR-sensor is inspired by a manual created by Seeedstudio [47] in which the Seeeduino V4.2 with a base shield was selected as the platform. Furthermore, the temperature sensor DS18B20 is connected to ground and a Digital pin on the Seeeduino communicating over a 1-wire bus. We solder a pull-up resistor of 4.7 KΩ at the back of the base shield to avoid unnecessary wiring, thus make the devices as compact as possible.

Since we want the sensing system to be wearable, we include wireless data transmission by a bluetooth module. A Grove Bluetooth Low Energy (BLE) V1 module is connected to the base shield by using one digital pin at the Seeeduino board for the receiver and another digital pin for the transmitter. Grove BLE V1 is a Low Energy Bluetooth module provided by Seeedstudio. The module has ATtention (AT) command support and can communicate with phones without pairing. A schematic of the BLE V1 is shown in 3.6.

For easy and compact connection, both the GSR sensor and the BLE belong to the Grove family which is a modular, standardized prototyping system. It provides a base shield that maps some of the Seeeduino pins to 4 point connectors. The modules provided by Grove contain equivalent 4 point connectors. The Grove system originates from Seeedstudio. Figure 3.7 shows the Seeeduino with a base shield and the connection of the different components.

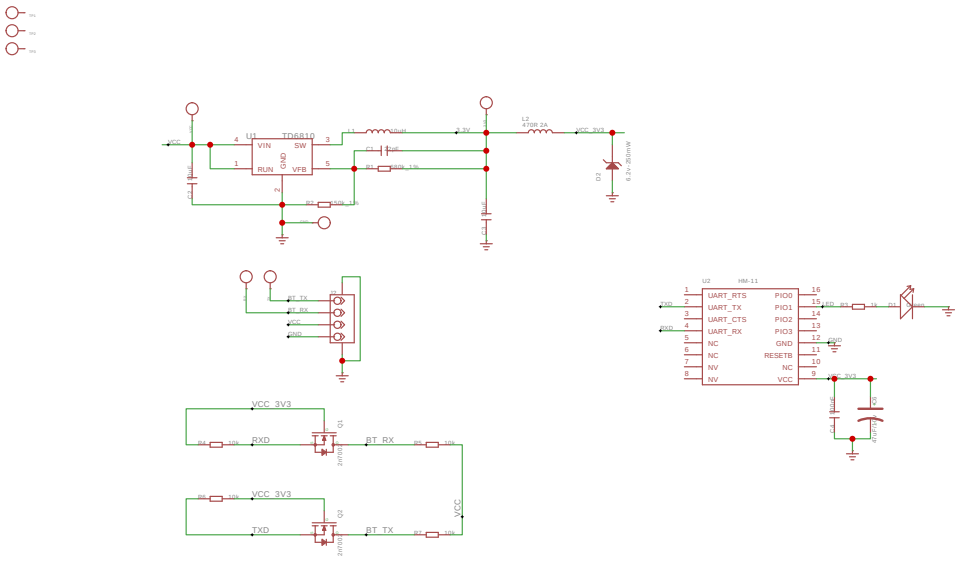


Figure 3.6: Schematic of the Grove BLE v1. Image courtesy of [3]

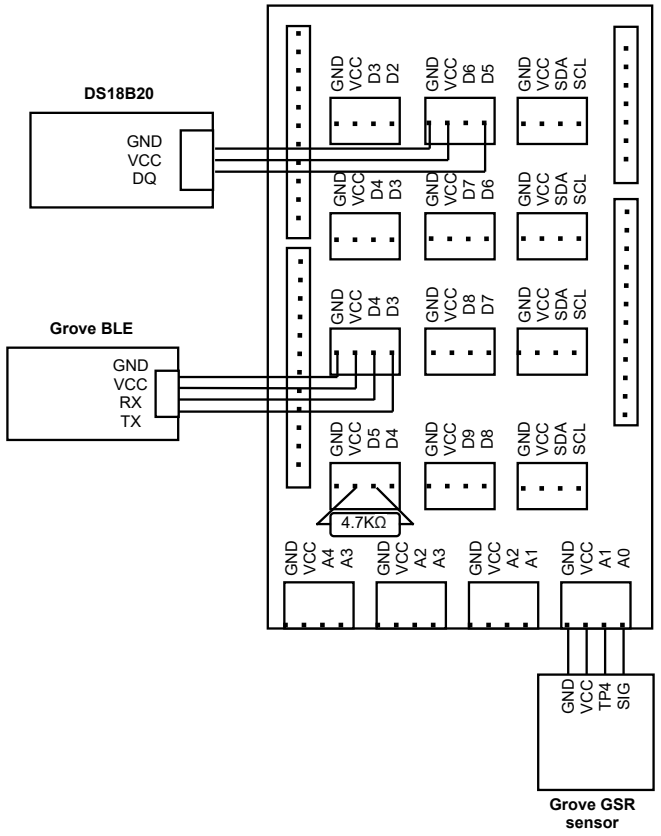


Figure 3.7: Seeeduino with base shield and connected modules.

3.1.3 An iOS Application: Health Data Collector

Our sensing system consists of two independent devices, one seeeduino-based sensing system with an attached GSR sensor and digital thermometer, and a Polar H10 pulse belt. This creates an issue of synchronizing the two devices. All the sensors have to start simultaneously or register a timestamp when the different sensors start tracking. There is also a need to adjust the tracking interval to be the exact time interval the participant watches the film excerpt.

Another issue is that the Polar H10 pulse belt is a commercial workout aid with encrypted information transmitted by BLE. Moreover, the mobile application provided by the Polar team themselves shows a summary of the workout rather than the raw, time series workout data, which is of interest to us.

Fortunately, the suppliers of the Polar H10 provide an Software Development Kit (SDK) for Android and iOS, which allows analysis of the detailed time series. We then

decide to develop an ad-hoc iOS application for automatically starting all of the sensors by the click of a button and save a timestamp with each sampling data.

The developed mobile application is named Health Data Collector (HDC) which will be used further in this thesis. The communication between the different modules is shown in figure 3.8.

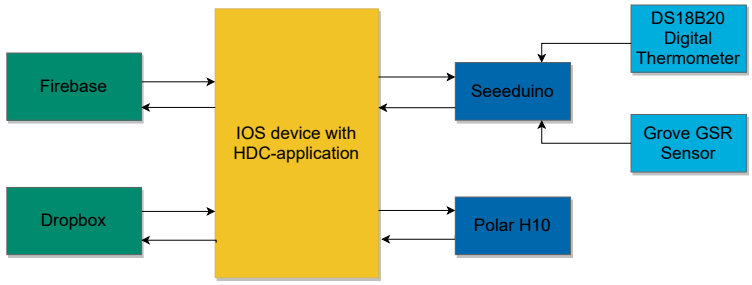
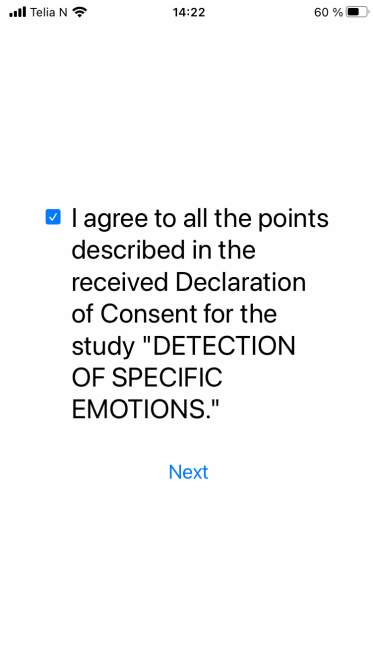


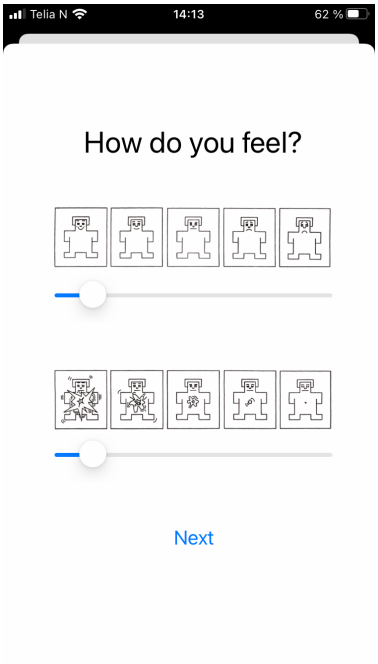
Figure 3.8: Component diagram of Health Data Collector(HDC) IOS application and communicating modules

Heart rate, RR-intervals, and ECG-signals are measured by the Polar H10 by using premade functions, and we use a swift framework called "Core Bluetooth" to communicate with the Seeeduino. All the collected health data are automatically saved in a personal Dropbox folder of the researcher when pushing a send button. For this achievement, we use SwiftyDropbox - a swift SDK for Dropbox. We also integrate a real-time database provided by Firebase to make it possible to conduct online analysis for future work.

Ease of use is a high priority when developing a mobile application. The HDC allows easy data collection of groups of people while leaving space for an extension of the project in the future. Figure 3.9 shows each view of the HDC throughout one recording session. As one can observe from the figure, both a consent of participation and a self-assessment inspired by the SAM described in section 2.1.2 are integrated into the app. Furthermore, the procedure of collecting data must happen in chronological order. The application ensures this order by activating and deactivating buttons. In fact, only one button in each view is active at a time so that the participant and the researcher know which button to push at any time. This includes the consent in the first view. The experiment cannot start without consent, in which the "next" button will not be activated before the check box is checked. To integrate the self-assessment is beneficial in many ways. First of all, it is compact, and there is no need to print out any papers. It will always appear in the correct order, so that the chronological order of the experiment is ensured. A desirable continuous scale is easy to present in the form of a slider, and the values are nicely saved in the dropbox folder with the collected sensor data. A complete manual explaining how to use the HDC, following the protocol used for this thesis, is found in Appendix C.


Telia N
14:22
60 %
I agree to all the points described in the received Declaration of Consent for the study "DETECTION OF SPECIFIC EMOTIONS."
Next

(a)


Telia N
14:13
62 %
How do you feel?
Next

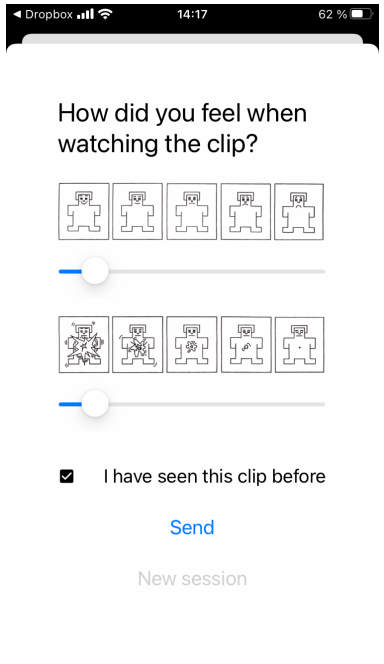
(b)


polarBleSdkTest...
16:01
51 %
Avbryt
Dropbox
HealthDataCollector ønsker å:
Rediger innhold i Dropbox-filer og -mapper, vis innhold i Dropbox-filer og -mapper og vis og rediger informasjon om Dropbox-filer og -mapper, kun innenfor mappen Apper › HealthDataCollector
Vis grunnleggende informasjon om Dropbox-kontoen, for eksempel brukernavn, e-post og land
Learn about permissions
Tillat
Bruk en annen konto

(c)


Telia N
13:16
4 %
Polar ID:
7895C82F
Clip ID:
TA
Start
Stop
Next

(d)



(e)

Figure 3.9: An overview of the different views of the HDC - application. (a) The participant has read an information sheet regarding the experiment. The experiment cannot proceed without their consent. (b) The participant declares how he/she feels by sliders and a SAM-inspired questionnaire. (c) The application asks for permission to edit and shows relevant content and information in dropbox. (d) The researcher types the ID of the Polar belt and the Clip ID, which is different for each clip excerpt and each task. The film clip is shown as the stop button turns blue. This means that all the sensors have started tracking. The researcher pushes stop when the excerpt ends. (e) The participant answers a questionnaire once again, this time about how they felt while watching the excerpt. If the the "New Session" button is pushed, a new session starts from view (b).

3.1.4 Selection of Elicitation Method

Literature shows a great variation in methods used to evoke emotions. A majority of related experiments have used either pictures, videos or music as elicitation methods. Such methods are beneficial as it gives the researcher control over the test situation and makes it possible to compare the experiments on the same basis, as opposed to elicitation methods in which the participants have a more active role such as games or letting the participant think and write about a situation in which a specific emotion was triggered.

In this thesis, we chose video excerpts for evoking emotions to control the standardization of the experiments and to evoke stronger emotions than either music and pictures alone may elicit. Furthermore, a limited number of film excerpts needed to be selected. This selection was made by the following three steps:

1. We selected a database containing carefully selected film excerpts evaluated by conducting a large survey.

2. We chose the excerpts in the database that were most suitable for our purpose and added them to a pool of candidate films.

3. We conducted a survey ourselves on a representative group of participants to evaluate the films in the pool of candidate films. Based on this survey we finally chose four excerpts that elicited emotions in each of the quadrants of Russel's Circumplex model in figure 2.1.

Selection of Candidate Film Excerpts

We endeavoured to select four film excerpts to evoke four distinct emotions; each one in a different quadrant of Russel's Circumplex model; and have a length of approximately 3-5 minutes. The search for four optimal film clips started in the literature from similar experiments; however, even though there exists literature describing similar experiments with film clips as the emotional elicitation method, very few studies specify exactly which film clips they use in the experiments.

Baveye et al. [7] list and compare 7 databases for affective content analysis and related applications. Just one of the databases in this list matches our criteria of being approximately 3-5 minutes long excerpts. That is the FilmStim database [40] which contains 70 film excerpts with duration 1-7 minutes.

There are some aspects of the FilmStim database that one should note. First, Schaefer et al. [40] decided the database's content based on 50 experts recalling and citing film excerpts in the different emotional categories. The database includes the ten most frequently cited excerpts in each category. Therefore, the majority of the film excerpts are extracted from a large amount of well known, old classics. This is not necessarily an issue, but we should take it into consideration. Having seen the entire film before may strengthen the emotions if the excerpts are extracted from a film in an overall equal emotional category or, on the other hand, it may give disturbing associations if the excerpts are extracted from a film in an overall different emotional category. It is also reason to believe that emotions are poorly triggered if one knows what will happen, especially for the emotional category of fear which is closely related to the fight-or-flight response. Furthermore, the technology is developing fast within the film industry, and the fact that the films are old may also be a critical issue for the fear category. The blood and the creatures in the newest horror movies look different and maybe more real than in the old horror movies. It may be a distraction if one starts to reflect upon the poor technology of the past. Second, the film excerpts are labeled by questionnaires from 364 undergraduate students with a predominance of women and a mean age of 19.6 from a Belgian, French-speaking university. It is well known that there are cultural differences in emotional reactions. The paper of Schaefer et al. [40] includes a quite homogeneous group in their studies. Social and cultural differences should be considered. Notice also that some excerpts are in French.

Hypothesis About the Labels in the FilmStim Database The film excerpts in the FilmStim database are labeled with the emotional categories amusement, sadness, anger, tenderness, disgust, fear, and neutral. On the basis that we want to detect one emotion in each of the quadrants in the valence-arousal plane we made the following hypothesis.

We assumed that the film excerpts in the FilmStim database could be placed in the four quadrants of the valence-arousal plane. Table 3.1 shows the hypothesis about the correspondence between the quadrants and the FilmStim-labels. We selected amusement in First Quadrant (Q1), tenderness in Second Quadrant (Q2), sadness in Third Quadrant (Q3), and disgust/fear in Fourth Quadrant (Q4). We filtered out the french excerpts and set the lower limit for the duration to be 2 minutes and 51 seconds. We also filtered out excerpts with a notable heterogeneous emotional mood.

Table 3.1: Hypothesis about which quadrants in Russell's Circumplex model the labels of the FilmStim database belongs to.

Quadrants	Labels
Q1	Arousal
Q2	Tenderness
Q3	Sadness
Q4	Fear, Anger, Disgust

However, the filtering lead to a poor selection of film excerpts for some of the categories. Thus, we added more modern excerpts to the fear category. Wanhui Wen and Huang [53] achieved an accuracy of 84% for classifying fear by using excerpts from the horror film "The Ring". The paper does not mention which scenes of the films are used, so three excerpts were evaluated, extracted, and added to the pool of candidate films. There was also a need for adding more material to the category of sadness. Gross and Levenson [22] evaluate different film excerpts for emotion elicitation. The ending scene from the film "The Champ" is highlighted as very efficient for eliciting sadness in which the excerpt resulted in a high and dominant rating in the emotional category of sadness. Furthermore, Rakshit et al. [37] used this excerpt for emotional elicitation and achieved 81% accuracy of recognizing sadness. Therefore we also added the ending scene from the "The Champ" to the pool of candidate films. The complete selection of candidate films for each emotional category is found in table 3.2.

Table 3.2: An overview of the candidate film excerpts divided into four categories.

Sadness	Amusement	Tenderness	Fear/Disgust
The Champ (Ending scene.)	There is Something about Mary(2) (Mary takes sperm from Ted's ear.)	Life is Beautiful (Reunited with mother.)	The silence of the lambs (Extraction of a butterfly's larva from a dead body's mouth)
City of angles (Maggie dies in Set's arms.)	There is Something about Mary(1) (Ted fights with the dog.)	The dead poet society (O Captain! my Captain!)	The Shining Jack pursues his wife with an axe.
	When Harry met Sally (Sally fakes orgasm)		The Ring(1) (Rachel has a scary dream)
			The Ring(2) (Rachel falls into a well)
			The Ring(3) (Noah dies)

Questionnaire for Choosing the Four Best Film Excerpts

In order to select the four optimal films from table 3.2, one questionnaire was made for each of the emotional categories of fear/disgust, sadness, amusement and tenderness. We invited 16 people to answer the questionnaires. In an attempt to reflect the test group that would later watch the final four selected films and contribute health data, the invited people included a mixed group of 8 females and 8 males, from 24 to 64 years old, in which the majority were in their 20s. All of the participants had the opportunity to decide for themselves which and how many questionnaires they wanted to answer.

The film excerpts were rated by the SAM, in which the participants rated their emotional state when watching the film excerpts by valence and arousal, both in the range of 1 to 10.

The mean score for each of the film excerpts is shown in table 3.3, and figure 3.10 shows the films placed in the valence/arousal plane based on the questionnaires with the axis crossing in (V,A) : (5.5,5.5). The quadrants of the plane are separated by the different colors, green, red, yellow and blue. The points in the plane are colored in the same color as the quadrant we assumed, from the hypothesis in paragraph 3.1.4, the points would belong to.

One can observe from figure 3.10 that some of the points reject the hypothesis in 3.1.4 in which the corresponding excerpts are evaluated to belong to a different quadrant.

It seems like there is especially great confusion among the green points that represent amusement in which all of them are placed in different quadrants. There is also worth noting that they are all placed close to the intersection between the axis which indicates that the excerpts did not have a great emotional impact on the participants. Notice also point 8 which is the only excerpt with the label disgust in the FilmStim database. According to the participants does this excerpt evoke too little arousal in order to belong to Q4.

Table 3.3: The results of the questionnaire for choosing four optimal film excerpts. represented by the mean value of the valence and the arousal.

Film excerpt	Mean Valence	Mean Arousal	Number of Answers
The Champ	3	4.67	6
City of angles	4.33	3.17	6
There is Something About Mary(1)	4.4	5	5
There is Something About Mary(2)	6	5.2	5
When Harry Met Sally	5.6	7	5
Life is Beautiful	7.8	5	5
The dead poet society	8.4	6	5
The Silence of Lambs	2.75	4.75	4
The Shining	3.5	6.75	4
The Ring(1)	2.75	7.5	4
The Ring(2)	2.75	7.25	4
The Ring(3)	2.75	6.75	4

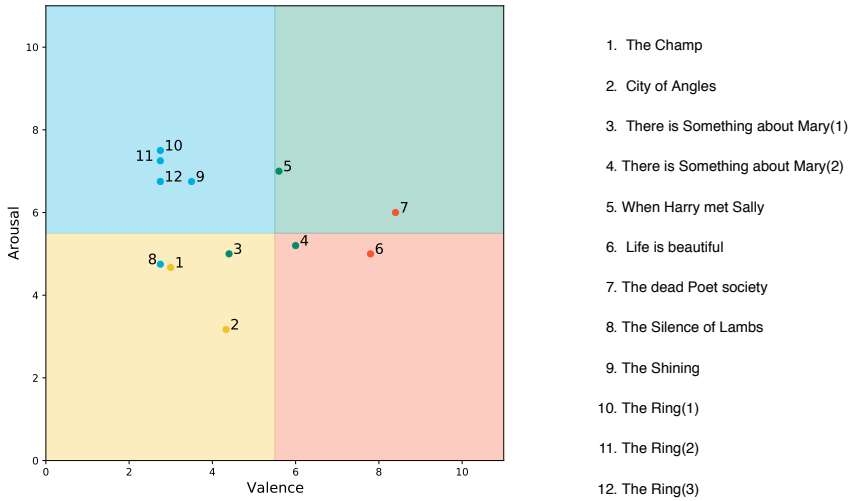


Figure 3.10: The values in table 3.3 plotted in a valence/arousal plane. Each colored dot must be placed inside the quadrant with corresponding color for the hypothesis in table 3.1 to be valid. The figure shows that, according to the survey, only some of the films meet the hypothesis.

Table 3.4 lists the final four selected excerpts. We chose “When Harry met Sally” for representing Q1 and “Life is Beautiful” for Q2 because those excerpts are the only ones belong to each of the corresponding quadrants. Furthermore for selecting the excerpts,

representing Q3 and Q4, we compared figure 3.10 with Russel's Circumplex model in figure 2.1 page 7. We chose the point within Q3 and Q4 in figure 3.10 that is placed closest to the point labeled "sad" and "afraid", respectively, in figure 2.1.

Figure 3.11 shows the individual validation of the four final selected films. One can observe from the bar chart that there are no consensus among the participants in the survey about which quadrant each of the four excerpts belongs to - except from the category *fear*.

For a simpler notation we will name the four films Film Amusement (FA), Film Tenderness (FT), Film Sad (FS), and Film Fear (FF).

Table 3.4: The final selected 4 excerpts representing one quadrant each.

Quadrant	Film excerpt	Simpler notation
Q1	When Harry met Sally	FA
Q2	Life is Beautiful	FT
Q3	The Champ	FS
Q4	The Ring(1)	FF

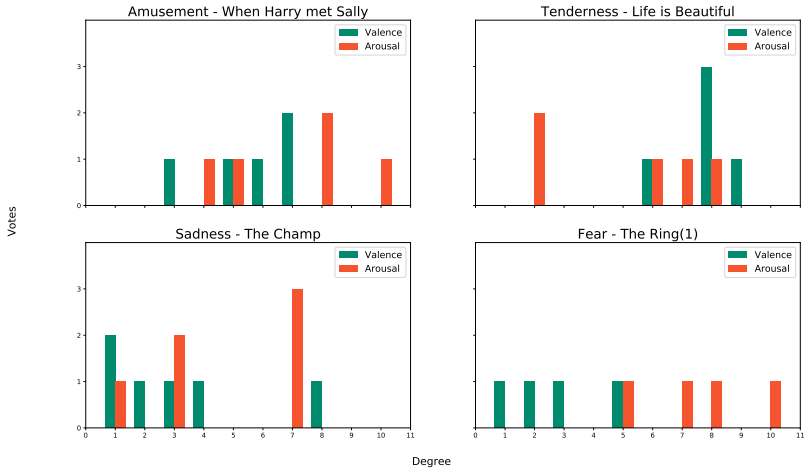


Figure 3.11: A bar chart for how the participants in the survey evaluated the different 4 selected excerpts. One can observe that there is no consensus about which quadrant the excerpts belong to, except The Ring(1) in which all the participants evaluated high arousal and low valence.

3.1.5 Experimental Protocol

When having selected the different sensors used for collecting the bio-data and chosen the method for evoking emotions. An experimental protocol was left to be defined before

conducting any experiments.

A total of 13 females and 11 males were asked to participate in the experiment. The age varied between 21 and 74 in which the majority of the participants were in their 20s. All participants were healthy without obesity problems or heart diseases. One experimental session lasted for approximately 50 minutes and included

- Calibration of the temperature sensor,
- Informed consent,
- Agreeing to participate in the study indicated by a checkbox in the application,
- Watching a non arousing and neutral film except to record a baseline,
- Watching the 4 selected film excerpts to trigger different emotions,
- Answering a SAM integrated in the application both before and after watching each film excerpt,
- Doing cognitive distraction tasks for 5 minutes between the emotional film excerpts to neutralize the emotions.

The participant watched the four emotion triggering film excerpts in a random order. [5, 24, 12, 21] gave inspiration for selecting the cognitive distraction tasks in which they have used maze solving, Sudoku, searching for specific words in a text, and copying geometric shapes for emotion mitigation in their studies. The participants completed the tasks also in a random order.

We selected the first 5 minutes of Attenborough's Planet Earth episode "Jungles" as the baseline excerpt.

A complete and detailed description of the protocol can be found in Appendix A. All of the researchers involved in the data collection followed this detailed protocol carefully.

3.2 General Data Protection Regulation: An Ethical Perspective on the Experiments Performed in this Project

The experiments conducted in association with this thesis come with an ethical responsibility. In line with the General Data Protection Regulation (GDPR), when the research involves human beings as research objects, one need a legal basis. The GDPR introduces informed consent as one alternative to such basis in which the participant is informed when and which data will be collected, processed, stored and shared. All direct and indirect identifiers have to be anonymized to ensure that a person can not be identified from the collected data.

Since the project involves collecting, processing and storing data, we have submitted our protocol and consent form to the Norwegian Center for Research Data (NSD) to be

approved for further experiments. We exaggerate that we are not collecting confidential and sensitive data. Thus, there is no combination of the obtained information can be traced back to the individuals. The informed consent used for the experiments is found in Appendix B.

3.3 Three Hypotheses Regarding the Labeling of the Data Samples

Section 2.1 describes how the subjective and the unconscious aspects of emotions make it hard to truly define emotions. For our study, these elements are essential for how we want to label the data samples. *Should we rely on the self-assessments trusting the participants to truly know what they are feeling and that emotions are highly subjective? or should we trust a statistical survey that gives an idea of what one should feel, thus using the labels from table 3.4?*

To investigate the issues that arise when defining emotions, as described in section 2.1, we state three hypotheses regarding labeling the data samples.

Hypothesis 1 The participants are completely aware of their own emotions: The labels are solely decided from the participants' self-assesment.

Hypothesis 2 The participants are partly aware of their own emotions: Only the samples where the questionnaire corresponds to the hypothesis in table 3.4 is used in the analysis.

Hypothesis 3 The unconscious aspect dominates the subjective aspect: The labels are completely decided from the hypothesis in table 3.4.

3.4 Data Analysis

After collecting as much data as possible, we can begin analyzing the samples. The data analysis is done by different steps including preprocessing of the data, feature extraction, feature engineering, dimension reduction by Principal Component Analysis (PCA) and classification by Support Vector Machines (SVM). The hypotheses described above are considered when we analyze the collected bio-signals in which each hypothesis is tested for each analysis that is being done.

3.4.1 Preprocessing

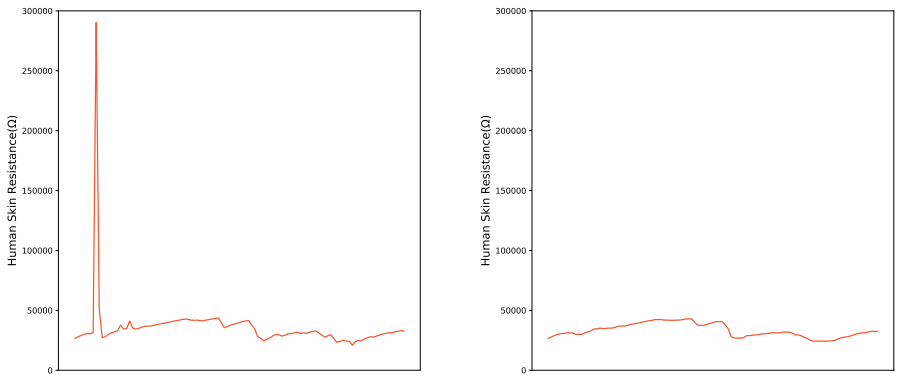
Before we apply data-driven models to our data, some preprocessing of the data is required. The preprocessing is done by the following steps.

Crop the Samples into Equal Time Lengths To ensure that the results are independent of the sample length, the samples generated from watching each of the emotional films (FA, FF, FS, FT) are cut to equal length. FA and FS have the shortest length so those are

used as a reference. The FF-samples are cut in the end because some of the participants said they felt more afraid at the beginning of the view, while the FT-samples are cut at the start in which some participants stated a calm and positive feeling at the end of the excerpt.

Conversion of Galvanic Skin Response Port Reading to Human Skin Resistance The port readings from the GSR sensor depends on the skin resistance, but it is not the actual value. The skin resistance is calculated by equation (3.1) page 25.

Filtering artifacts in GSR signal We filter the GSR signals by applying a moving-median filter described in section 2.5.1 in which rapid-transient artifacts were discovered in some of the GSR signals. Figure 3.12 shows a GSR signal before and after applying a moving-median filter.



(a) **(b)**

Figure 3.12: (a) Raw GSR signal, (b) Filtered GSR signal by applying a moving median filter.

3.4.2 Compensating for the Individuality of Biosignals

From chapter 2 we know that physiological signals are highly individual and may be affected by a large number of factors. There is a need for somehow compensating for those individualities. For that purpose, the following two methods are proposed.

Detecting a Change in Emotion by Removing the Baseline

A film excerpt with no purpose of triggering any specific emotion is shown at the start of each experiment. The mean of the different signals captured while watching this excerpt is subtracted from each signal captured when watching a film excerpt to trigger a specific emotion. Thus we produce signals which reflect differences from a start emotion. That

means that instead of detecting if the participant is sad or happy, we detect if the participant is sadder or happier. The main idea behind this method is that by removing a baseline, we remove information that is particular to a participant. However, the method can also be beneficial for seeing that, in the self-assessment, a participant's response to an excerpt has a tendency to depend on their previous responses.

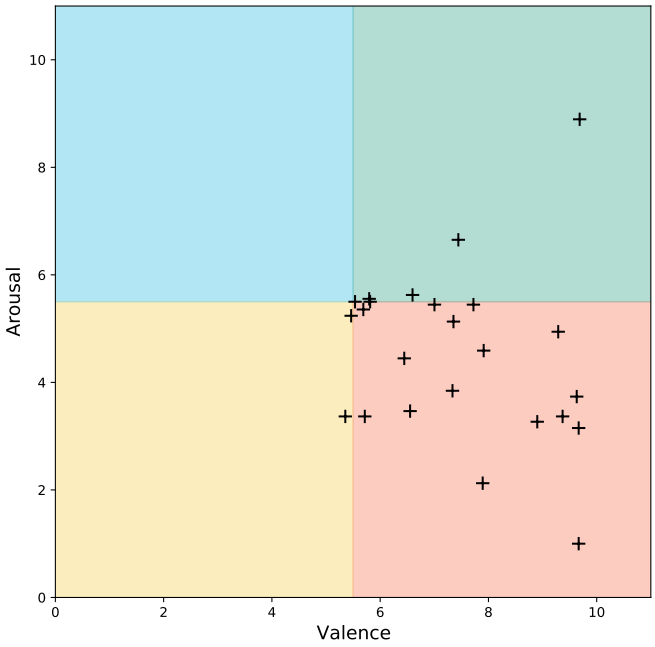
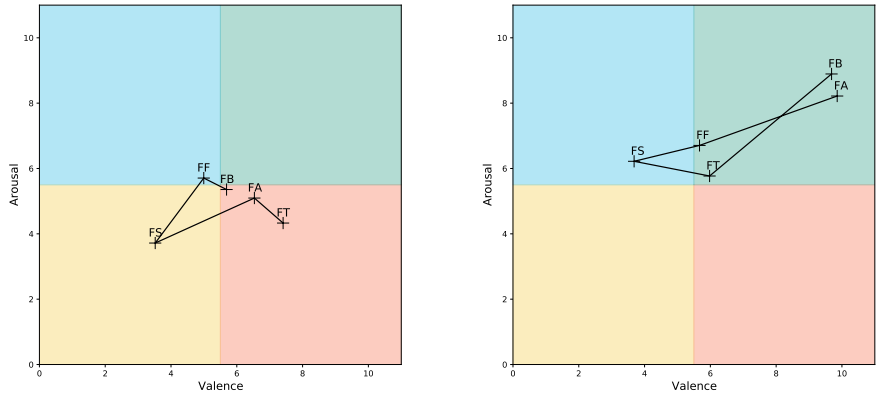


Figure 3.13: Valence/Arousal plane. The black crosses represents the self-assessments for the participant watching the baseline excerpt. As one can observe from the figure, most of the participants felt calm and positive in the when watching this excerpt.



(a) **(b)**

Figure 3.14: The figure shows the emotional journey of two participants. The black crosses represent the self-assessments for each excerpt triggering amusement (FA - When Harry met Sally), fear (FF - The Ring(1)), Sadness (FS - The Champ), and Tenderness (FT - Life is beautiful). The assessment for the baseline is marked as Film Baseline (FB). (a) The self-assessments almost correspond to the hypothesis in table 3.4 except the assessment of FA, which is placed in Q2 (the quadrant of tenderness). (b) Only the FA's assessment matches the hypothesis in table 3.4. However, the relation between the assessments of the excerpts triggering emotions is matching the hypothesis perfectly. A small translation of the entire plane along the valence axis plus a small translation along the arousal axis the points would be in the correct place according to the hypothesis in table 3.4

Detecting an Emotion without Removing the Baseline

In the second method, we are detecting emotions in their entirety. Instead of removing a baseline effect, we compensate for individuality by performing a min-max scaling of each sample. The min and the max are defined from the entire sample set from one participant. Thus the values of each sample depend on the entire emotional journey. The Power Spectral Density (PSD) of the RR-intervals are also scaled in this manner.

3.4.3 Feature Selection

The feature selection is inspired by Wanhui Wen and Huang [53], Rakshit et al. [37], Udovičić et al. [50], and Priyanka Das and Tibarewala [36]. For the RR-intervals, we extract features in both the frequency domain and the time domain. Note that we include both short-time variability indices by including jumps between two successive intervals and long-term variability indices in which statistical features such as standard deviation are calculated from the entire sample. As explained in section 2.3.2, extracting features from the RR-signals' power spectrum has shown good results in similar experiments. The frequency band is split into three bands: Very Low Frequency (VLF) 0-0.04 Hz, Low Frequency (LF) 0.04-0.15 Hz, and High Frequency (HF) 0.15-0.5 Hz as shown in figure 3.16. The PSD is found by Welch's method (described in section 2.5.3) with a window size

a quarter of the total signal length and a Hann window. Table 3.5 contains the extracted features from the different signals with an explanation.

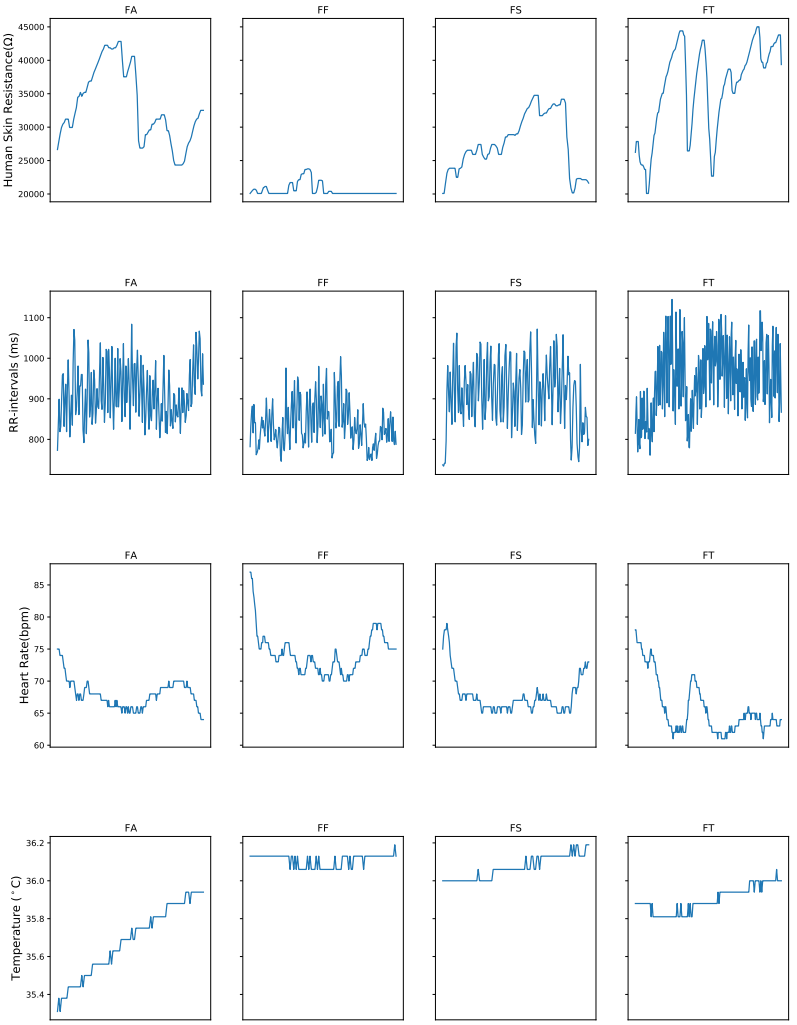


Figure 3.15: Collected biosignals for one participant, the GSR signal is converted to human skin resistance and filtered by a moving-median filter (see section 2.5.1). The rest of the graphs show raw signals.

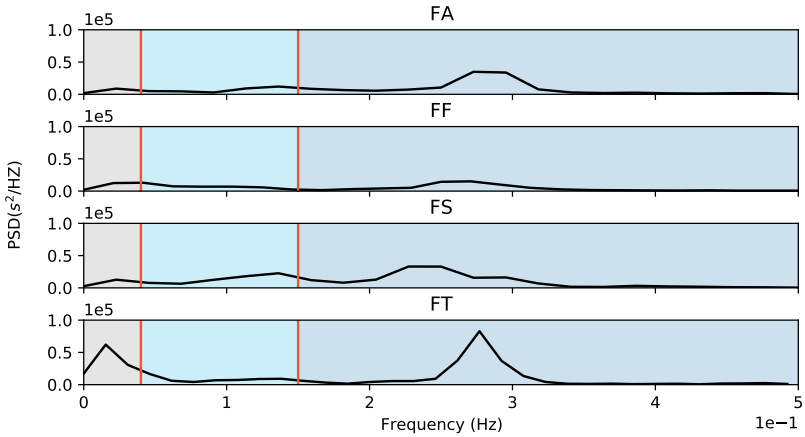


Figure 3.16: Power Spectral Density (PSD) found by Welch's method (described in section 2.5.3) corresponing to the RR-intervals in figure 3.15. The spectra are split into three frequency bands. The gray area shows frequencies in the VLF band, the light blue area shows frequencies in the LF band, and the dark blue area shows frequencies in the HF band.

Table 3.5: An overview of the features used for classification with description.

Type of signal	Feature	Description
Heart Rate	meanHR	Mean value of heart rates
	stdHR	Standard deviation heart rates
RR (Time domain)	meanRR	Mean value of RR intervals
	medianRR	Median value of RR intervals
	stdRR	Standard deviation of RR intervals
	semRR	Standard error of the mean value of RR intervals.
	RRfracjumps	Percentage of successive RR differing by greater or equal to 50 ms
	rmsRR	Root mean square of successive RR differences
RR (Frequency domain)	maxVLF	Highest power in VLF band
	maxLF	Highest power in LF band
	maxHF	Highest power in HF band
	aVLF	Raw area of VLF band
	aLF	Raw area of LF band
	aHF	Raw area HF band
	aTOT	Total raw area of VLF, LF and HF bands
	pVLF	Relative VLF area w.r.t total area
	pLF	Relative LF area w.r.t total area
	pHF	Relative HF area w.r.t total area
	nLF	Normalized LF area (aLF/(aLF+aHF))
	nHF	Normalized HF area (aHF/(aLF+aHF))
	LFHF	Ratio of LF and HF areas (aLF/aHF)
GSR	meanGSR	Mean value of GSR signal
	stdGSR	Standard deviation of GSR signal
	slopeGSR	The slope of a linear regression line of the GSR signal.
	maxvalGSR	The maximum value of the GSR signal.
	minvalGSR	The minimum value of the GSR signal.
	kurtosisGSR	The kurtosis of the GSR signal.
	skewnessGSR	The skewness of the GSR signal
GSR derivatives	GSRFDsqrtMean	Root mean square of first derivative GSR signal
	GSRFDmean	Mean value of first derivative GSR signal.
	GSRSDsqrtMean	Root mean square of second derivative GSR signal
	GSRSDmean	Mean value of the second derivative GSR signal.
Temperature	meanT	Mean value of the temperatures.
	stdT	Standard deviation of the temperatures.
	minvalT	Minimum value of the temperatures.
	maxvalT	Max value of the temperatures.
	modeT	The mode of the temperatures.

3.4.4 Dimensionality Reduction by Principal Component Analysis and Classification by Support Vector Machines

For classification, we use SVM, described in section 2.7 with an RBF kernel, which is a frequently classification method for similar problems [57, 56, 36, 18, 50, 37, 43]. Furthermore, PCA (see section 2.6) is applied for dimension reduction creating a simpler model that is easier to visualize.

It is a well known fact, e.g., see [1, 4], that when building supervised learning models, one ideally wants to minimize the bias-variance trade off, i.e., have a model that captures the variances in the training data and generalizes for unseen data. Unfortunately, this has been proven to be a difficult task. There is in fact a bias-variance trade-off, which means that one must settle down for a balance that minimizes the total error of bias and variance.

One commonly searches for an optimal model complexity in order to minimize the bias-variance trade off. For that purpose, we start with a simple model including only one Principal Component (PC), then increase the number of components and perform a leave-one-out cross validation each time we add a new component.

In addition to classification based on all the quadrants Q1 (Amusement), Q2 (Fear), Q3 (Sadness) and Q4 (Tenderness), we apply SVM to the following two classification problems.

1. A binary classification problem in which we consider a pair of the quadrants, i.e., Q1/Q2, Q1/Q3, Q1/Q4, Q2/Q3, Q2/Q4 and Q3/Q4.

2. Considering the two half planes of the valence-arousal plane first divided by the arousal axis (high valence against low valence), and then divided by the valence axis (high arousal against low arousal)

3.4.5 Feature Engineering

One of the most time consuming steps when working with classification algorithms is to find an optimal set of features that will give the best accuracy. Table 3.5 shows a base set of features. There is, however, a need for investigation of another combination of features that would increase the performance. For this purpose we perform the following.

1. Remove some features based on the loading for the principal components. If we observe a drop in the performance when adding certain PCs increasing the complexity of our model, we investigate the loading for the PC that causes the drop and remove features with high influence for the variation of that particular PC. In the same manner, we also investigate peaks, and only keep the features with a high influence on the variation for that particular PC.

2. Generate new features by multiplying all possible pairs of features in the original set. In that way we include interactions between the features.

Chapter 4

Results

This chapter shortly presents the performance and results from the analysis described in the previous chapter. Two methods compensating for subjectivity in bio-signals have been tested. That is removing a baseline and thus detect a change in the emotional state, and applying a min-max normalization detecting emotions in their entirety.

Furthermore, as there is little agreement on how to define emotions, we have presented 3 hypotheses in the methodology about how to label the data samples. These hypotheses are:

1. Base the labels solely on the self-assessment
2. Only include samples that corresponds to the statistical survey in section 3.1.4.
3. Label the data solely based on the statistical survey.

The chapter summarizes the results for each hypothesis and each method compensating for subjectivity in bio-signals in a different subsection. For the sake of brevity we present what we believe are the five most significant results that have been obtained through the proposed numerical analyses.

4.1 Result 1: The Optimal Model Complexity Varies from each of the Hypotheses

The SVM performance corresponding to the number of Principal Components included in our model is depicted in figure 4.1 and figure 4.2. As one can observe, there is no common optimal number of Principal Components for the three hypothesis. In most of the cases, the optimal model complexity is found by including a few of the Principal Components, but this is not consistent.

4.1.1 Detecting Emotions in their Entirety

Figure 4.1 shows that when we detect the emotions in their entirety the optimal model complexity is found by including a single Principal Component for hypothesis 2 and four principal components for hypothesis 3. The performance of the classification based on hypothesis 1 shows an increasing trend, and the optimal model complexity is in fact found by including 19 principal components.

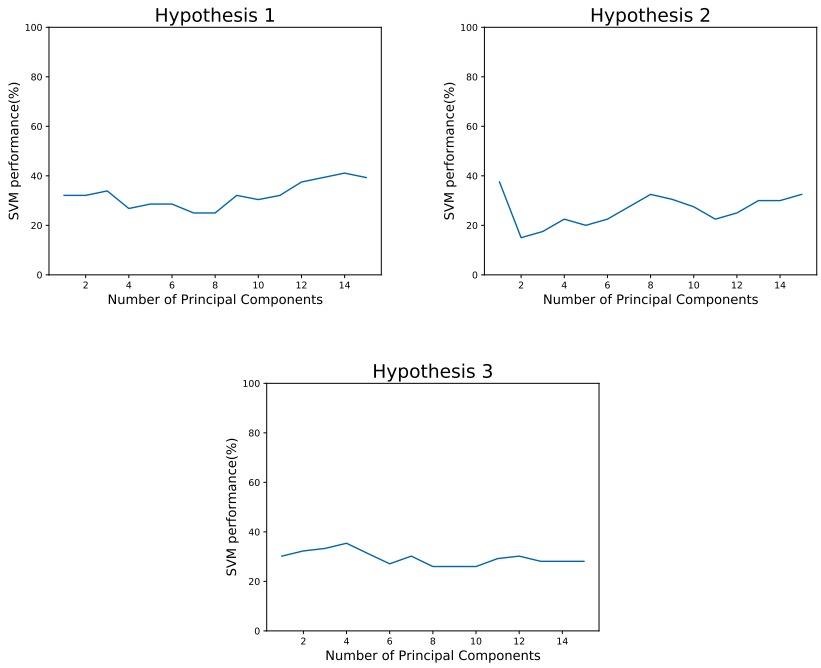


Figure 4.1: The figure shows the SVM performance relative to the model complexity for each of the three hypotheses formulated in section 3.3 when detecting emotions in their entirety. As one may observe, only a few Principal Components constitute the optimal model complexity for hypothesis 3 and hypothesis 2, while there is a need for including a high number of Principal Components to obtain the optimal model complexity for hypothesis 1.

4.1.2 Detecting Changes in an Emotional States

Figure 4.2 shows that when we detect changes in emotional states by removing a baseline, the optimal model complexity is found by including only two Principal Components for hypothesis 1 and hypothesis 3, while ten Principal Components are added to achieve an optimal model complexity for hypothesis 2.

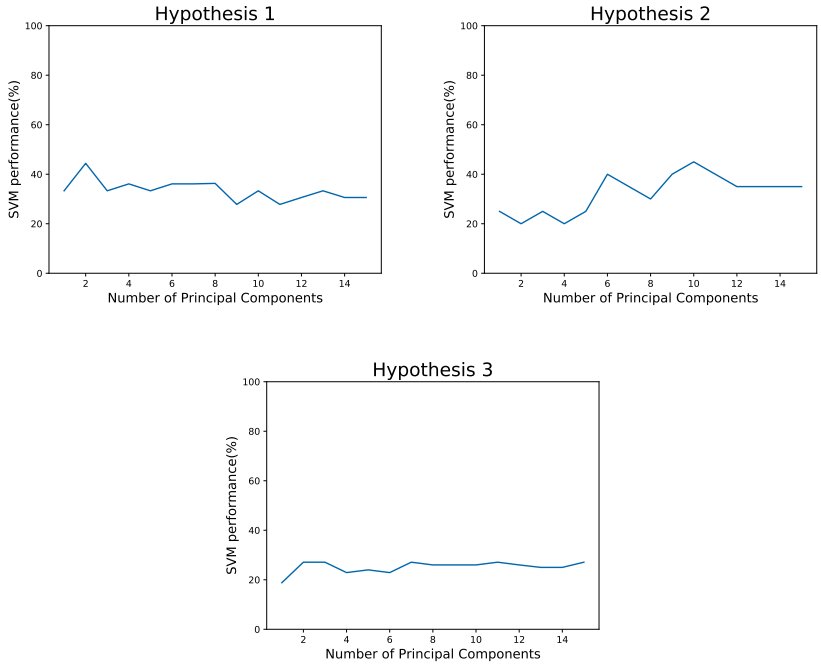


Figure 4.2: The figure shows the SVM performance relative to the model complexity for each of the three hypotheses formulated in section 3.3 when detecting a change in an emotional state. As one can observe, only a few Principal Components constitute the optimal model complexity for hypothesis 1 and hypothesis 3, while there is a need for including a high number of Principal Components to obtain the optimal model complexity for hypothesis 2.

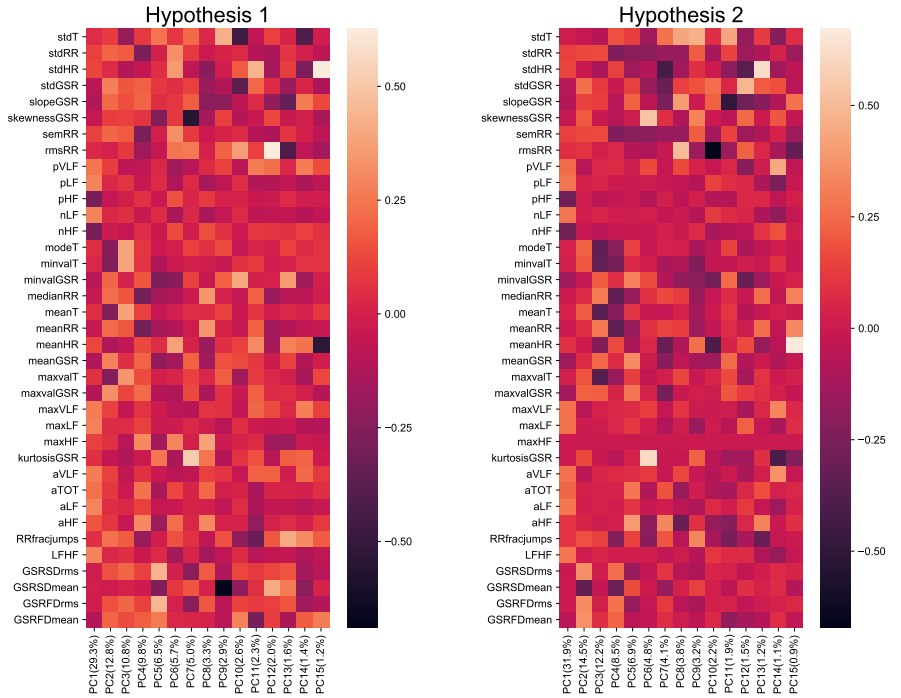
4.2 Result 2: The Feature Selection Process Influences the Principal Component and Feature Engineering Results

Generating new features by multiplication did not seem to increase the SVM performance for any of the scenarios. However, by studying the plots in figure 4.2 and figure 4.1, we increased the performance for some of the hypothesis by removing some of the features that seemed to contribute to the drops in figure 4.2 and figure 4.1. The increased performance for the different hypothesis and the corresponding combination of features is shown in table 4.1 and table 4.2.

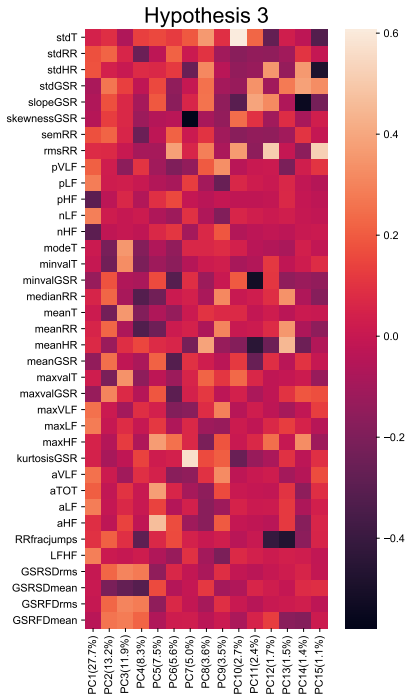
4.2.1 Detecting Emotions in their Entirety

Table 4.2 shows the performance by using some selected features from table 3.5 when detecting an emotion in its entirety. We could not find a better performance for hypothesis

3 by using another combination of features than the original feature set. For hypothesis 1, and hypothesis 2 however, we were able to increase the performance to 50% accuracy. To obtain this accuracy for hypothesis 1 we only removed five features. Eleven features were removed for hypothesis 2. As one can observe, all of the features considering the body temperature except stdT were removed.



(a) **(b)**



(c)

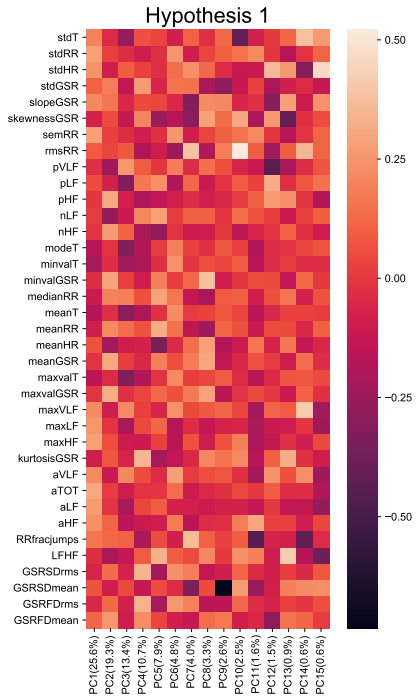
Figure 4.3: The heatmaps shows the loadings for PC1 to PC15 for each of the hypotheses when detecting emotions in their entirety. The loading plots for each hypothesis is different in which the sample set for hypothesis 1 and hypothesis 2 are balanced. However, one can observe some similarities between the loading plots of the different hypothesis, e.g. pHF and nHF seem to have a great influence on the variation.

Table 4.1: We were able to increase the performance by a removing some of the features in table 3.5. As one may observe we removed 16 features for hypothesis 2. Notice that sdtT is the only remaining feature from the temperature signal. The features stdRR, semRR, maxHF, kurtosisGSR and skewnessGSR were removed from the original feature set when considering hypothesis 1.

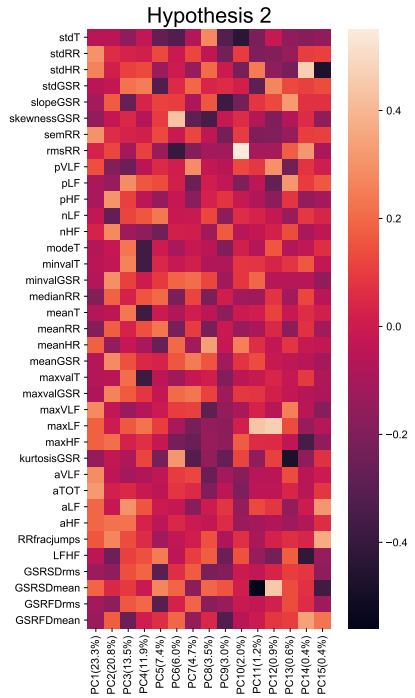
	SVM performance All features	**SVM performance New feature set**	**New feature set**	**Principal Components**
Hypothesis 1	42.9%	50.0%	GSRFDmean, GSRFDrms, GSRSDmean, GSRSDrms, LFHF, RRfracjumps, aHF, aLF, aTOT, aVLF, maxLF, maxVLF, maxvalGSR, maxvalT, meanGSR, meanHR, meanRR, meanT,medianRR, minvalGSR, minvalT, modeT, nHF, nLF, pHF, pLF, pVLF, rmsRR, slopeGSR, stdGSR, stdHR, stdT	PC1 - PC17
Hypothesis 2	37.5%	50.0%	GSRSDmean, GSRSDrms, LFHF, aLF, aTOT, aVLF, kurtosisGSR, maxLF, maxVLF, meanGSR, minvalGSR, nHF, nLF, pHF, pLF, pVLF, rmsRR, skewnessGSR, slopeGSR, stdHR, stdT	PC1, PC2, PC3
Hypothesis 3	35.4%	35.4%	No change	PC1,PC2,PC3,PC4

4.2.2 Detecting Changes in Emotional States

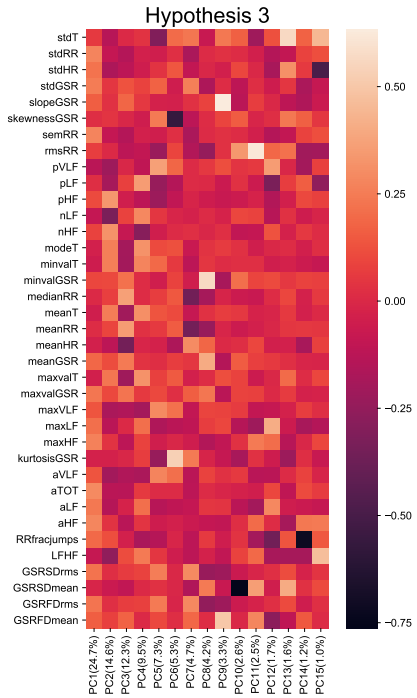
Table 4.2 shows the performance by using some selected features from table 3.5 when detecting a change in the emotional state by removing a baseline. We could not find a better performance for hypothesis 1 and hypothesis 2 by using another combination of features than the original feature set. For hypothesis 3, however, we were able to increase the performance by 9.4% by using a combination of just five features. Notice that those features are extracted from the GSR signal and the PSD of the RR-intervals - mostly from the high frequency band.


Hypothesis 1
stdT
stdRR
stdHR
stdGSR
slopeGSR
skewnessGSR
semRR
rmsRR
pVLF
pLF
pHF
nLF
nHF
modeT
minvalT
minvalGSR
medianRR
meanT
meanRR
meanHR
meanGSR
maxvalT
maxvalGSR
maxVLF
maxLF
maxHF
kurtosisGSR
aVLF
aTOT
aLF
aHF
RRfracjumps
LFHF
GSRSDrms
GSRSDmean
GSRFDrms
GSRFDmean
PC1(25.6%)
PC2(19.3%)
PC3(13.4%)
PC4(10.7%)
PC5(7.9%)
PC6(4.8%)
PC7(4.0%)
PC8(3.3%)
PC9(2.6%)
PC10(2.5%)
PC11(1.6%)
PC12(1.5%)
PC13(0.9%)
PC14(0.6%)
PC15(0.6%)
0.50
0.25
0.00
−0.25
−0.50

(a)


Hypothesis 2
stdT
stdRR
stdHR
stdGSR
slopeGSR
skewnessGSR
semRR
rmsRR
pVLF
pLF
pHF
nLF
nHF
modeT
minvalT
minvalGSR
medianRR
meanT
meanRR
meanHR
meanGSR
maxvalT
maxvalGSR
maxVLF
maxLF
maxHF
kurtosisGSR
aVLF
aTOT
aLF
aHF
RRfracjumps
LFHF
GSRSDrms
GSRSDmean
GSRFDrms
GSRFDmean
PC1(23.3%)
PC2(20.8%)
PC3(13.5%)
PC4(11.9%)
PC5(7.4%)
PC6(6.0%)
PC7(4.7%)
PC8(3.5%)
PC9(3.0%)
PC10(2.0%)
PC11(1.2%)
PC12(0.9%)
PC13(0.6%)
PC14(0.4%)
PC15(0.4%)
0.4
0.2
0.0
−0.2
−0.4

(b)



(c)

Figure 4.4: The heatmaps shows the loadings for PC1 to PC15 for each of the hypotheses when detecting a change in an emotional state.

Table 4.2: We were unable to improve the performance of hypothesis 1 and 2. The performance considering hypothesis 3 was increased by only including three features extracted from the GSR signal and the HF- and LF band of the RR-intervals.

	SVM performance All features	**SVM performance New feature set**	**New feature set**	**Principal Components**
Hypothesis 1	44.4%	44.4%	No change	PC1, PC2
Hypothesis 2	45.0%	45.0%	No change	PC1-PC10
Hypothesis 3	27.1%	36.5%	pHF, nHF, aHF, maxvalGSR, slopeGSR, LFHF	PC1, PC2, PC3

4.3 Result 3: The Emotion *Fear* Seems Easiest to Separate

In order to investigate if some of the classes are easier to separate, we conducted a binary classification with pairs of classes. Hypothesis 2, when detecting an emotion in its entirety, seems to have the best performance for this classification problem with an accuracy above 70% for all combinations. One can observe from table 4.3 that fear/sadness seems to be the easiest to separate and that all of the combinations that include fear have the highest accuracy for this hypothesis.

4.3.1 Detect emotions in their entirety

Table 4.3 shows the accuracy for a binary classification problem considering all possible pairs of the four classes when detecting emotions in their entirety. Hypothesis 2 produced an overall better performance than the two other hypothesis. The placement of the data samples is depicted by a score plot in figure 4.5 when testing hypothesis 2 for the three classification problems that produced the best accuracy. That is all the problems including the emotion *fear*.

Table 4.3: Performance considering binary classification problems when detecting emotions in their entirety. Hypothesis 2 seem to produce the overall highest accuracy.

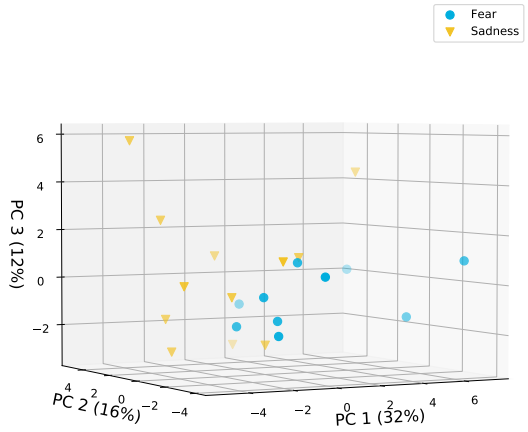
	Hypothesis 1	**Hypothesis 2**	**Hypothesis 3**
Sadness/Fear	78.6%	85.0%	70.8%
Amusement/Fear	82.1%	80.0%	62.5%
Tenderness/Fear	67.9%	80.0%	60.4%
Sadness/Amusement	69.5%	79.0%	70.8%
Sadness/Tenderness	62.5%	75.0%	58.3%
Tenderness/Amusement	74.0%	70.0%	58.3%

4.3.2 Detect Changes in Emotional states

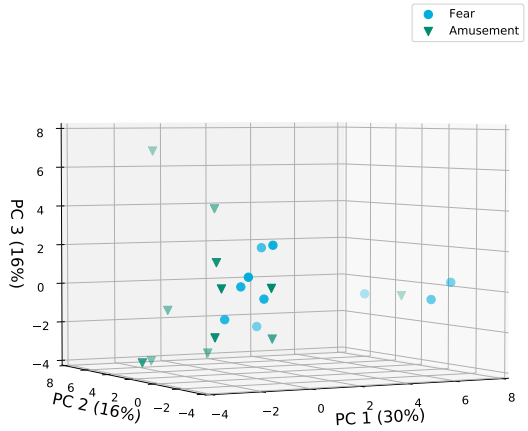
Table 4.4 shows the accuracy for a binary classification problem considering all possible pairs of the four classes when detecting changes in emotional states. None of the classification problems or hypothesis stands out to generate a superior result. Furthermore, table 4.4 shows an overall lower performance than table 4.3.

Table 4.4: Performance considering binary classification problems when detecting a change in the emotional state. Hypothesis 2 seem to produce the overall highest accuracy

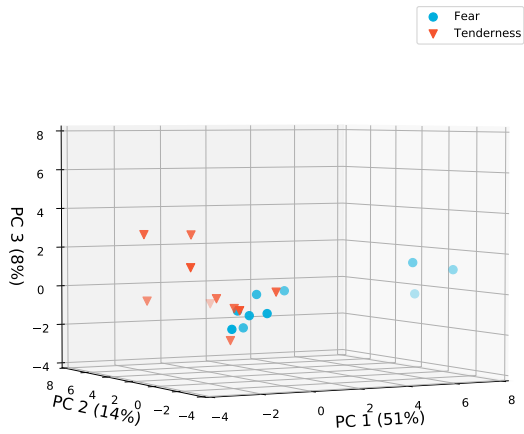
	Hypothesis 1	**Hypothesis 2**	**Hypothesis 3**
Sadness/Fear	52.8%	50.0%	64.6%
Amusement/Fear	54.8%	63.6%	66.7%
Tenderness/Fear	77.8%	58.3%	70.8%
Sadness/Amusement	58.3%	68.2%	58.3%
Sadness/Tenderness	72.2%	50.0%	58.3%
Tenderness/Amusement	72.2%	66.7%	52.1%


Fear
Sadness
PC 3 (12%)
PC 2 (16%)
PC 1 (32%)

(a)


Fear
Amusement
PC 3 (16%)
PC 2 (16%)
PC 1 (30%)

(b)



(c)

Figure 4.5: The three plots show how samples labeled as fear with each of the three other classes is placed in the space spanned by three principal components. The figures show that the case of detecting emotions in their entirety and the samples are labeled by hypothesis 2. Note that the data sets are balanced for classification (a) An accuracy of 85% was obtained. (b) An accuracy of 80% was obtained. (c) An accuracy 80% was obtained.

4.4 Result 4: Performing Detection on the Valence Scale or Arousal Scale has No Significant Difference

Table 4.5 and table 4.6 show the accuracy when detecting high valence/low valence and high arousal/low arousal of the data samples. The tables show that there is no significant difference in the performance when classifying on the valence scale compared to classifying on the arousal scale.

4.4.1 Detecting Emotions in their Entirety

Table 4.5 shows the performance for the two binary classification problems classifying high valence against low valence, and high arousal against low arousal when detecting emotions in their entirety. Table 4.5 shows that there are no significant difference in the performance of the two problems. figure 4.7 and figure 4.6 shows the placement of the

data points for high valence against low valence by hypothesis 2 and high arousal against low arousal by hypothesis 3 by a score plot.

	Hypothesis 1	**Hypothesis 2**	**Hypothesis 3**
HV/LV	65.0%	65.9%	57.4%
HA/LA	65.9%	65.9%	66.7%

Table 4.5: Performance when detecting high valence/low valence and high arousal/low arousal when detecting emotions in their entirety.

4.4.2 Detecting Changes in an Emotional States

Table 4.5 shows the performance for the two binary classification problems classifying high valence against low valence, and high arousal against low arousal when detecting changes in emotional states. As for the case of detecting emotions in their entirety, table 4.5 do not show a significant difference in the performances of the two problems.

	Hypothesis 1	**Hypothesis 2**	**Hypothesis 3**
HV/LV	58.5%	67.6%	63.5%
HA/LA	62.5%	58.8%	57.3%

Table 4.6: Performance when detecting high valence/low valence and high arousal/low arousal when detecting a change in an emotional state.

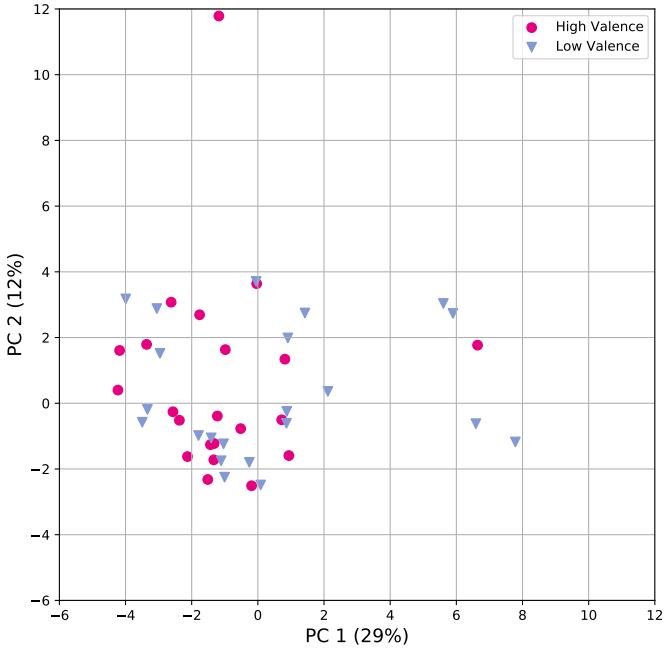


Figure 4.6: Detection of high valence/low valence when detecting emotions in their entirety by hypothesis 2. Note that an accuracy of 65.9% is obtained by including 6 Principal Components.

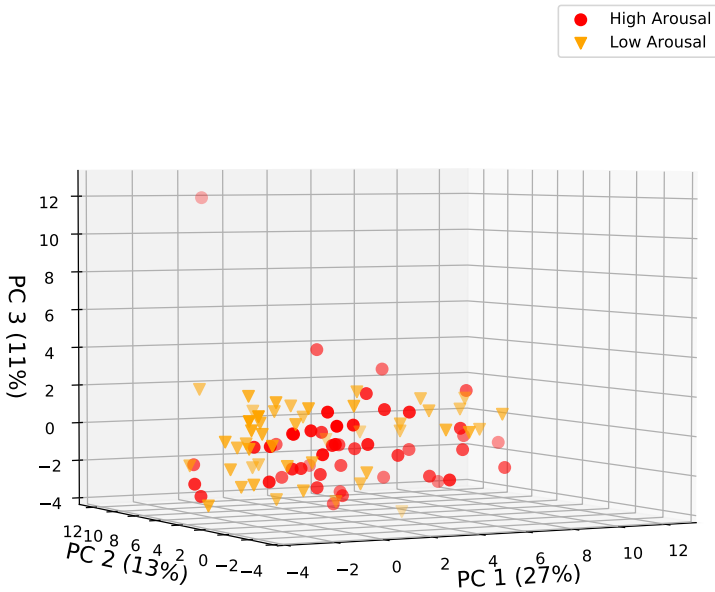


Figure 4.7: Detection of high valence/low valence when detecting emotions in their entirety by hypothesis 3. The accuracy of 66.7% is obtained by including 3 Principal Components.

4.5 Result 5: Detecting Emotions in Their Entirety by Hypothesis 2 Seems to be the Best Approach

By comparing the results of all the classification problems presented above it seems that predicting emotions in their entirety by hypothesis 2 is the better approach. Figure 4.8 shows a Principal Component Analysis including four classes by hypothesis 2 when detecting emotions in their entirety.

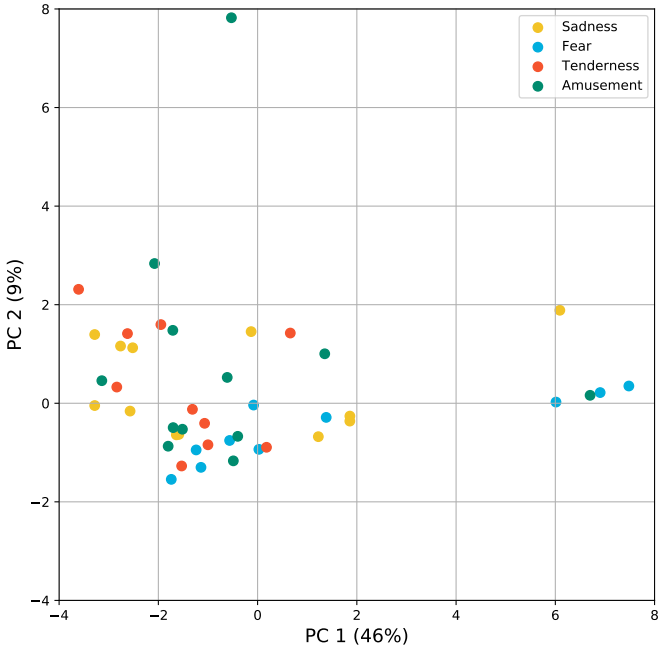


Figure 4.8: The placement of the samples by hypothesis 2 in a plane spanned by two PCs when detecting the emotions in their entirety. 50.0% accuracy is obtained by using 3 PCs. Note that the set is balanced when applying SVM.

Chapter 5

Discussion

It is difficult to compare our results with previous work, since the emotions and the number of detected feelings vary from study to study. Furthermore, many previous experiments have collected a lot of data from few participants, which reduces subjectivity in the data sets. In this thesis we investigated one of the more difficult tasks, i.e., detecting four distinct emotional states by conducting one experiment for each participant.

In chapter 4 we saw that we reached an accuracy of 50% when including all four classes. This is better than a random guess of 25% accuracy, but the results do not seem to reach the accuracy of similar experiments mentioned in the literature that use larger and more expensive sensors. For example, for the sake of comparison, Yoo et al. [55] proposed a neural network-based method for detecting the four quadrants in the valence/arousal plane, and reported an accuracy of 80.2%. However, multiple experiments were done by including just six subjects.

Udovičić et al. [50] reports instead results from using a compact, low-cost Shimmer3 sensor for the experiments; this paper shows quite similar results to the results obtained in this thesis, with an accuracy of 67% and 68.3% for detecting high valence/low valence, high arousal/low arousal, respectively when applying a multi-user model.

Interestingly, but maybe not surprisingly, the results show a relatively high accuracy when detecting fear. We know from theory that the *fight-or-flight response* is highly connected to the bio-signals that we have collected in our trials (while, instead, there is much more uncertainty about how the Autonomic Nervous System (ANS) reflects other emotions). Notice, however, that the excerpt of the emotion fear (FF) does not include any obvious jump scares and that just a few of the self-assessments for the excerpt of fear (FF) did contain extreme values for valence and arousal. Being able to detect fear can be useful for various applications, especially in the context of mental disorders; an application area that we discussed in the introduction). For example, it is well known that anxiety disorders and fear are directly connected[10], in the same way stress and fear are also highly connected, and the existing scientific models suggest that both are caused by survival instincts [2]. Studies show that the responses of stress and fear controlled by the ANS have similarities, and as mentioned in the introduction, stress over time can trigger various mental

disorders.

On the other hand, the results we summarized in table 4.3 seem to reflect a difficulty in detecting tenderness. The reason for such a result may be due to a suboptimal selection of the used film excerpt (FT). After the experiment, many of the participants said that they were confused about the excerpt. In this case, we may have failed to show excerpts that were triggering a homogeneous emotional mood. Even though there are positive music and pleasant dialogues, the action takes place in a war situation that may have a negative effect on the mood.

5.1 A Larger Data Set Would Provide a Greater Scientific Foundation

We should also report that we faced a common problem, i.e., not having access to plenty of data samples – something we believe to have affected our final experimental results greatly. More precisely, 24 participants is not entirely sufficient to base any statistically rigorous scientific inference claim. Moreover, the collected dataset were further reduced when testing hypothesis 2, in which only a fraction of the self-assessments correspond to the hypothesis formulated in table 3.4. Besides, the dataset is balanced when applying the classification model. When detecting a change in the emotions, the dataset was originally highly unbalanced. As one can observe from figure 3.13 most of the baseline assessment points are placed in Q2. It seems that “Our Planet” (FB) and“Life is Beautiful” (FT) somehow triggers the same emotions. Thus samples labeled as “Tenderness” by the self-assessments are underrepresented in the data set. When detecting a changes in emotional states based on hypothesis 2, the data set only contains 5 samples for each class.

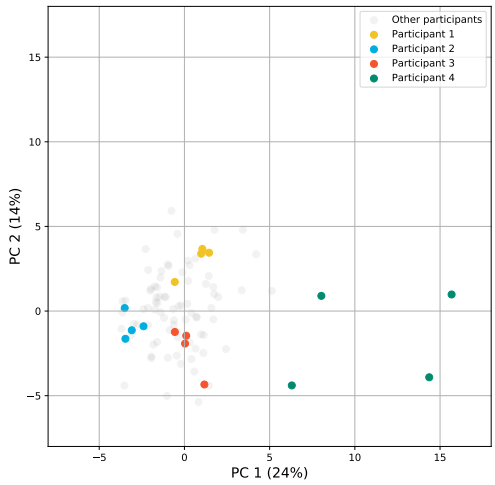
With a larger time window for research, one may also consider collecting more data from each participant to conduct user-specific detection.

5.2 On Compensating for Subjectivity by Removing a Baseline

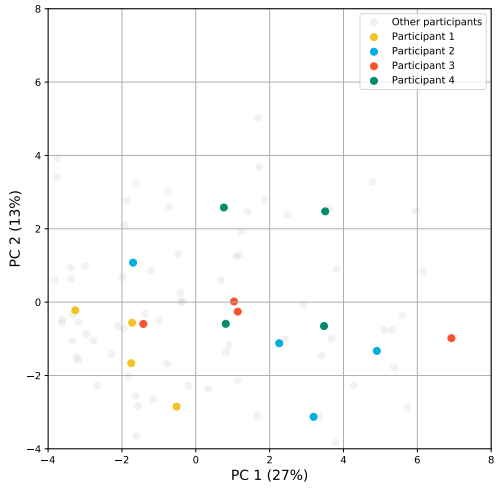
From the results summarized in chapter 4, it seems that to detect emotions in their entirety by applying a min-max normalization for the sake of compensating for subjectivity gives an overall slightly better performance than detecting a change in an emotional state by removing a baseline. One important issue when removing a baseline, which should be investigated in future work (and potentially changed), is about showing the baseline excerpt (FB) at the start of the experiment. Even though most of the participants answered by the self-assessment that they were calm and positive when watching the baseline excerpt (FB), we have seen a tendency in which the heart rate (BPM) is at its highest just at the beginning of the experiment, possibly caused by an expectation of what will happen.

Figure 5.1 shows how the samples of four randomly selected participants are placed in a plane spanned by two principal components calculated from the feature set in table 3.5.

When we detect both a change in an emotional state and when we detect emotions in their entirety, we fail to remove the subjectivity to some extent in which it is easy to spot clusters in the data set. Nevertheless, we see that the participants are much more intertwined when we normalize by a min-max normalization than when we remove a baseline for suppressing subjectivity.



(a)



(b)

Figure 5.1: The placement of the samples FA, FF, FS and FT for four different participants. Note that if the samples of one specific participant form a cluster, then this indicates that we fail to some extent to remove the effects of subjectivity from those samples. (a) In this plot a baseline is removed to compensate for subjectivity (b) Here instead a min-max normalization method is applied to compensate for subjectivity effects. This seems a better strategy, since empirically the clusters seem to do not form as clearly as in the other case.

Determining good methods for compensating for subjectivity seems crucial for boosting the performance of the classification methods. As mentioned before, studies using more samples from fewer subjects have shown a better performance than we have achieved in this experiment. For example, Gu et al. [23] conducted a subject-independent study, with the same emotional targets as we proposed in this thesis, achieving an accuracy of 50.3% for four classes, which is almost exactly what we have achieved. Furthermore, Gu et al. [23] proposed a novel biometric signature-based system by separating the data set into groups based on biometric characteristics, thus transforming subject-independent problems into multiple subject-dependent problems. Using this method, Gu et al. [23] were able to increase the performance from 50.3% to 90.7%

5.3 The Hypothesis Considering Labeling the Data Samples

Testing the different hypothesis described in section 3.3 is interesting in which differences in the results may reveal some information about emotional consciousness. It does not seem like unconscious emotions dominate the results in which hypothesis 3 produces a significantly lower accuracy than hypothesis 1 and hypothesis 2. Hypothesis 2 shows a slightly better overall performance, especially when considering pairs of the classes. Nevertheless, it is an assertion on insufficient grounds to say that this difference is due to emotional unconsciousness. It may just as well be due to other factors such as differences in interpretation of the SAM.

5.4 Issues Regarding the Experimental Protocol

Finally, we should discuss the experimental protocol and the test situation. We have already mentioned a tendency in which the participant feels a bit tense or excited at the beginning of the experiment. Furthermore, as mentioned in the background, chapter 2, a test situation is an arena for regulating their emotions. They know that their emotions will be tested and can prepare for it.

A majority of the participants mentioned after watching the film excerpt of sadness (FS) that they didn't feel so sad because they got thrown into a situation without knowing the background story and without building up a relation to the characters. However, there is a trade-off between showing long excerpts and short excerpts, in which the long excerpts will provide a more heterogeneous emotional spectrum. A solution could be to show complete films and let participants mark time slots where they have strong emotions. This would however be time-consuming.

Furthermore, many self-assessments show small variations in the valence- and arousal scale for one participant, indicating that the excerpts did not trigger strong emotions, despite the strong scientific basis for selecting the particular excerpts. Other, more non-traditional methods for triggering specific emotions proposed in the literature allow the

participants to play a more active role in the experiment. For example, letting the participant think and write about a situation that triggered a specific emotion [31]. We believe that such methods should be considered for further work, and compared against what we have been measuring in our experimental campaign.

Chapter 6

Conclusion and Future work

We have in this thesis designed, developed, and tested a low cost, wearable sensing system that can be used for measuring pulse, electrodermal activity, and skin temperature. Furthermore we have designed, organized and executed a small-scale data collection trial on volunteers, and used the collected data as inputs to a set of different statistical methods to detect specific emotions. In doing so we have tested different hypotheses and strategies for analyzing the signals. We have seen that none of the hypotheses and methods produced an accuracy sufficiently high to be used in a commercial device when including all four classes. Nevertheless, we have seen a tendency in which the emotion *fear* is easier to detect than the other emotions, with our algorithms showing a final accuracy above 80% for all of the binary classification problems that included detecting fear.

It is however difficult to compare our results with others from similar experiments reported in the literature, given that there is a great variation in experimental factors such as the physiological signals as indicators, emotions to trigger, and number of included subjects. A wide range of choices also makes it difficult to track where one should make changes. The following section will propose some changes that can be done for further experiments in addition to future extensions of the project.

6.1 Future Work

6.1.1 Improved Methods to Compensate for Subjectivity

We have mentioned in several places that one of the greatest challenges of detecting emotions across individuals is the presence of subjectivity in the bio-signals collected and used for the classification purposes. We have seen that we were able to remove this subjectivity to some extent by applying a min-max normalization method. There are however reasons to believe that other methods, such as the method developed by Gu et al. [23], may increase the performance of the SVM.

6.1.2 Collect More Data Samples

Due to time constraints and limited access to volunteers, a relatively small amount of samples have been collected. It would be interesting to explore how the classifiers' boundaries evolve with the addition of further data, and how fast detection performance increase (or plateau) by extending the dataset. Furthermore, it would be interesting to collect more samples for each individual, building single-user individualized models to be compared with multi-user 'average' ones.

6.1.3 Emotion Elicitation Methods

Many of the participants were skeptical about the experiment. They thought the excerpts were too short for building up emotions, and that it was challenging for several of them to be thrown into a film excerpt without knowing its context. Little research has been done on methods in which participants have a more active role in the experiment. Elicitation methods in which the participants play some game, for example are put into a Virtual Reality (VR), or think and write about experienced situations that trigger specific emotions may better reflect real-life situations and are interesting to explore.

6.1.4 Improved Classification Methods

We have restricted ourselves in using classification methods that have been commonly used in previous, similar experiments – more precisely, Support Vector Machines (SVM) with Radial Basis Fuction (RBF) kernels. However, other methods have shown good results as well [53, 55]. Applying different methods for this project is left to be explored.

6.1.5 Expansion of the Ancillary IT Infrastructure Supporting the Project

There are unlimited possibilities for aiding the project through opportune IT infrastructure. Some of the possibilities are already arranged, in which we have included communication with a real time database in the application for online emotion detection. This topic has shown to be very challenging and sophisticated tools are left to develop[44].

Furthermore, even though the sensing system we have developed can be battery powered and is relatively easily wearable, one should explore the possibility to integrate all the sensors in a simple wristband for future production of commercial devices.

Bibliography

[1] (2012). Bias/variance tradeoff. https://ocw.mit.edu/courses/sloan-school-of-management/15-097-prediction-machine-learning-and-statistics-spring-2012/lecture-notes/MIT15_097S12_lec04.pdf.

[2] (2020). Understanding the stress response. https://www.health.harvard.edu/staying-healthy/understanding-the-stress-response.

[3] (Cited 2021). Grove ble v1. https://wiki.seeedstudio.com/Grove-BLE_v1/.

[4] Avati, A. (Cited 2021). Bias-variance analysis theory. http://cs229.stanford.edu/summer2020/BiasVarianceAnalysis.pdf.

[5] Baddeley, A., Banse, R., Huang, Y.-M., and Page, M. (2012). Working memory and emotion: Detecting the hedonic detector. *Journal of Cognitive Psychology*, 24(1):6–16.

[6] Barrett, L. F. (2018). *How emotions are made*.

[7] Baveye, Y., Dellandréa, E., Chamaret, C., and Chen, L. (2015). Liris-accede: A video database for affective content analysis. *IEEE Transactions on Affective Computing*, 6(1):43–55.

[8] Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, COLT '92, page 144–152, New York, NY, USA. Association for Computing Machinery.

[9] Bradley, M. M. and Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential.

[10] Buhr, K. and Dugas, M. J. (2009). The role of fear of anxiety and intolerance of uncertainty in worry: An experimental manipulation. *Behaviour Research and Therapy*, 47(3):215–223.

[11] Carl, J. R., Soskin, D. P., Kerns, C., and Barlow, D. H. (2013). Positive emotion regulation in emotional disorders: A theoretical review. *Clinical Psychology Review*, 33(3):343 – 360.

[12] Christianson, S., Loftus, E., Hoffman, H., and Loftus, G. (1991). Eye fixations and memory for emotional events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17:693–701.

[13] Cortes, C. and Vladimir, V. (1995). Support-vector networks.

[14] Darwin, C. (1872/1998). *The expression of the emotions in man and animals*.

[15] Distrelec, E. (Cited January 2021). Elfa distrelec. `https://www.elfadistrelec.no`.

[16] Dr Jason J Braithwaite, Dr Derrick G Watson, R. J. and Rowe, M. (2019). A guide for analysing electrodermal activity(eda) skin conductance responses(scrs) for psychological experiments. `https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/psychology/ABriefGuideforAnalysingElectrodermalActivityv4.pdf`.

[17] Ekman, P. (1970). Universial facial expression of emotions. *California mental health research digest*, 8.

[18] Gaetano Valenza, Luca Citi, A. L. E. P. S. and Barbieri, R. (2014). Revealing real-time emotional responses: a personalized assessment based on heartbeat dynamics. `https://www.nature.com/articles/srep04998`.

[19] Garcia-Ceja, E., Riegler, M., Jakobsen, P., Torresen, J., Nordgreen, T., Oedegaard, K. J., and Fasmer, O. B. (2018). Motor activity based classification of depression in unipolar and bipolar patients. In *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*, pages 316–321.

[20] Giakoumis, D., Tzovaras, D., Moustakas, K., and Hassapis, G. (2011). Automatic recognition of boredom in video games using novel biosignal moment-based features. *Affective Computing, IEEE Transactions on*, 2:119 – 133.

[21] Grimshaw, G., Kranz, L., Carmel, D., Moody, R., and Devue, C. (2017). Contrasting reactive and proactive control of emotional distraction. *Emotion*, 18.

[22] Gross, J. J. and Levenson, R. W. (1995). Emotion elicitation using films. `https://bpl.berkeley.edu/docs/48-Emotion%20Elicitation95.pdf`.

[23] Gu, Y., Tan, S., Wong, K., Ho, M. R., and Qu, L. (2010). A biometric signature based system for improved emotion recognition using physiological responses from multiple subjects. In *2010 8th IEEE International Conference on Industrial Informatics*, pages 61–66.

[24] Hoppitt, L., Mathews, A., Yiend, J., and Mackintosh, B. (2010). Cognitive bias modification: The critical role of active training in modifying emotional responses. *Behavior Therapy*, 41(1):73 – 81.

[25] integrated, M. (2019). Ds18b20. https://datasheets.maximintegrated.com/en/ds/DS18B20.pdf.

[26] John F. Kihlstrom, S. M., Tobias, B. A., and Tobis, I. P. (2000). *Cognition and Emotion*.

[27] Kim, J. and André, E. (2008). Emotion recognition based on physiological changes in music listening. https://ieeexplore.ieee.org/abstract/document/4441720.

[28] Lewis, M., Haviland-Jones, J. M., and Barrett, L. F. (2008). *Handbook of emotions*.

[29] Mazefsky, C. A., Herrington, J., Siegel, M., Scarpa, A., Maddox, B. B., Scahill, L., and White, S. W. (2013). The role of emotion regulation in autism spectrum disorder. *Journal of the American Academy of Child Adolescent Psychiatry*, 52(7):679 – 688.

[30] Md.Yassin, F. and Chin, s. n. (2019). Analysis of heart rate and body temperature from the wireless monitoring system using arduino. *Journal of Physics: Conference Series*, 1358:012041.

[31] Mills, C. and D'Mello, S. (2014). On the validity of the autobiographical emotional memory task for emotion induction. *PLOS ONE*, 9(4):1–10.

[32] Olfson, M., Druss, B. G., and Marcus, S. C. (2015). Trends in mental health care among children and adolescents. *New England Journal of Medicine*, 372(21):2029–2038. PMID: 25992747.

[33] Oy, P. E. (2019). Polar h10 heart rate system. https://www.polar.com/sites/default/files/static/science/white-papers/polar-h10-heart-rate-sensor-white-paper.pdf.

[34] Polanczyk, G. V., Salum, G. A., Sugaya, L. S., Caye, A., and Rohde, L. A. (2015). Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of Child Psychology and Psychiatry*, 56(3):345–365.

[35] Polar (Cited January 2021). Polar h10. https://www.polar.com/uk-en/products/accessories/polar_h10_heart_rate_sensor.

[36] Priyanka Das, A. K. and Tibarewala, D. (2016). Emotion recognition employing ecg and gsr signals as markers of ans. https://ieeexplore.ieee.org/abstract/document/7746134.

[37] Rakshit, R., Reddy, V. R., and Deshpande, P. (2016). Emotion detection and recognition using hrv features derived from photoplethysmogram signals. In *Proceedings of the 2nd Workshop on Emotion Representations and Modelling for Companion Systems*, ERM4CT '16, New York, NY, USA. Association for Computing Machinery.

[38] Russell, J. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178.

[39] SAYKRS, B. (1973). Analysis of heart rate variability. *Ergonomics*, 16(1):17–32. PMID: 4702060.

[40] Schaefer, A., Nils, F., Sanchez, X., and Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition Emotion - COGNITION EMOTION*, 24.

[41] Schwarts, J., Gibb, W., and T, T. (1991). Aging effects on heart rate variation. *Ergonomics*.

[42] Seo, Y.-S. and Huh, J.-H. (2019). Automatic emotion-based music classification for supporting intelligent iot applications. *Electronics*, 8:164.

[43] Setyohadi, D., Kusrohmaniah, S., Gunawan, S., Pranowo, P., and Prabuwono, A. S. (2018). Galvanic skin response data classification for emotion detection. *International Journal of Electrical and Computer Engineering*, 8:4004–4014.

[44] Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., Xu, X., and Yang, X. (2018). A review of emotion recognition using physiological signals. *Sensors*, 18(7).

[45] Soares, A., Pinheiro, A., Costa, A., Frade, S., Comesaña, M., and Pureza, R. (2013). Affective auditory stimuli: Adaptation of the international affective digitized sounds (iads-2) for european portuguese. *Behavior research methods*, 45.

[46] Solomon, Jr, O. M. (1991). Psd computations using welch's method. [power spectral density (psd)].

[47] Studio, S. (cited January 2021). Grove-gsr sensor. `https://wiki.seeedstudio.com/Grove-GSR_Sensor/`.

[48] Suveg, C. and Zeman, J. (2004). Emotion regulation in children with anxiety disorders. *Journal of Clinical Child & Adolescent Psychology*, 33(4):750–759. PMID: 15498742.

[49] Tipping, M. E. and Bishop, C. M. (1999). Mixtures of probabilistic principal component analysers. `http://www.miketipping.com/papers.html`.

[50] Udovičić, G., erek, J., Russo, M., and Sikora, M. (2017). Wearable emotion recognition system based on gsr and ppg signals. In *Proceedings of the 2nd International Workshop on Multimedia for Personal Health and Health Care*, MMHealth '17, page 53–59, New York, NY, USA. Association for Computing Machinery.

[51] van der Meer, L., van't Wout, M., and Aleman, A. (2009). Emotion regulation strategies in patients with schizophrenia. *Psychiatry Research*, 170(2):108 – 113.

[52] van Ravenswaaij-Arts, C. M. A., Kollee, L. A., and Hopman, J. C. (1993). Heart rate variability. *Annals of Internal Medicine*, 118(6):436–447. PMID: 8439119.

[53] Wanhui Wen, Guangyan Liu, N. C. J. W. P. S. and Huang, W. (2014). Emotion recognition based on multi-variant correlation of physiological signals. `https://ieeexplore.ieee.org/document/6825835`.

[54] Wen, W., Liu, G., Cheng, N., Wei, J., Shangguan, P., and Huang, W. (2014). Emotion recognition based on multi-variant correlation of physiological signals. https://ieeexplore.ieee.org/document/6825835.

[55] Yoo, S., Lee, C., Park, Y., Kim, N., Lee, B., and Jeong, K. (2005). Neural network based emotion estimation using heart rate variability and skin resistance. volume 3610, pages 818–824.

[56] Yu-Liang Hsu, Jeen-Shing Wang, W.-C. C. C.-H. H. (2017). Eeg-based emotion recognition in music listening. https://ieeexplore.ieee.org/abstract/document/8219396.

[57] Yuan-Pin Lin, Chin-Hong Wang, T.-P. J. T.-L. W. S.-K. J. J.-R. D. and Chen, J.-H. (2010). Eeg-based emotion recognition in music listening. https://ieeexplore.ieee.org/abstract/document/5458075.

Appendices

Appendix A

Experimental protocol

A.1 Introduction

The aim of this experiment is to collect health data using different types of sensors in order to do further processing and investigate the possibilities of emotion detection using data driven methods.

For this purpose, 24 subjects will participate in an experiment in which they will watch different film clips evoking different emotions.

A.2 Sensors and devices

The participants are equipped with sensors that monitors pulse, skin temperature and galvanic skin response. To be able to activate all the sensors at the same time, the iOS application ”health-data-collector” has been developed.

A.3 Film clips and material

4 emotional film clips, representing the 4 emotions fear, amusement, tenderness and sadness, are selected from the database FilmStim, and also from the films The ring and The champ which are all used in published, related papers such as [40, 54, 37].

To be able to decide a baseline, a neutral film clip is shown in the beginning of the experiment. For this purpose 5 minutes of an episode of David Attenborough's Planet Earth is used.

Emotion mitigation is required before deciding a baseline showing an Attenborough clip. For this purpose, the participant is performing a cognitive distraction task for 5 minutes which is randomly selected from the four activities Sudoku, geometric-shapes-copying, maze solving, find all *a*'s. All of the tasks have been used in published emotional experiments [5, 24, 12, 21].

A.3.1 Maze solving

- File: mazes_cognitiveTasks.pdf
- What to do: draw a path from the green arrow to the red arrow without crossing any of the ”walls”.

A.3.2 Sudoku

- File: sudoku_cognitiveTasks.pdf
- What to do: fill the number 1-9 in all cells so that each number is not repeated within any of the 3×3 squares or in any row.

A.3.3 Geometric-shapes-copying

- File: geometricShapes_cognitiveTasks.pdf
- What to do: draw the geometric shapes in the file on a blank sheet.

A.3.4 Find all *a*'s

- File: Dummytext_cognitiveTasks.pdf
- What to do: mark all *a*'s in the dummy text.

A.4 Self assessment

The participant is answering a small questionnaire before and after watching a film clip. The questionnaire is based on the self-assessment manikin (SAM) which is a standardized non-verbal pictorial assessment. The participant rates the valence and the arousal of his/her emotions through a set of SAM-pictures and sliders questionnaires that are integrated in the health-data-collector app. After watching the film clip the participant is also asked if he/she has seen the film clip before.

A.5 Protocol

A.5.1 Setting up the environment before the participant enters the room

1. The researcher downloads a copy of the ”information about the participant” file on her/his laptop, and opens that file for editing it.
2. The researcher opens the following web links, each showing an excerpt of an emotional film, in a browser in her/his laptop:
   (a) Baseline

(b) Tenderness

(c) Amusement

(d) Sadness

(e) Fear

This means that the browser will have 5 tabs open;

3. The researcher generates a randomized order of the list of excerpts "2b, 2c, 2d, 2e" (NOT the baseline) by:

   (a) opening `https://www.random.org/lists/`

   (b) copying the text "2b 2c 2d 2e" in the website, in the box indicated there (note that for doing this it will be necessary to put each element of that list on a separate line)

   (c) press the button "randomize". This will generate a randomized version of the list, and in the following we will refer to this randomized order with "`randomOrder`(2b-2c-2d-2e)"

4. The researcher organizes the tabs in his/her browser by following the randomized order above, i.e., as 2a - `randomOrder`(2b-2c-2d-2e), so that the emotional film excerpts will be shown to the participant in the generated random order. Note that the baseline clip 2a will *always* be shown first.

5. The researcher writes down the order of these clips in the "information about the participant" file

6. The researcher prints and organizes the list of cognitive distraction tasks by:

   (a) printing the consent form.

   (b) printing each of the following tasks

      i. the "maze solving" task described in Section A.3.1

      ii. the "sudoku" task described in Section A.3.2

      iii. the "geometric-shapes-copying" task described in Section A.3.3

      iv. the "find all *a*'s" task described in Section A.3.4

   (c) generating a random order for these cognitive tasks by using again the website `https://www.random.org/lists/` to shuffle the list 6(b)i, 6(b)ii, 6(b)iii, 6(b)iv. This means randomizing which tasks will be given to the participant between the various emotional film excerpts listed above (remember that also that ones have been shuffled in a random order).

7. The researcher writes down the order of these cognitive tasks in the "information about the participant" file.

8. The researcher sets up the hardware, i.e.,

   (a) if he/she has not done it yet, reads the manual for the application.

(b) Writes down the ID of the polar belt that will be used in the experiment somewhere (indeed this number is going to be used several times, thus better to have it at hand at anytime)

(c) turns on the iPhone, its Bluetooth, and its wifi connection (verifying that it is indeed connected to the Internet)

A.5.2 Making the participant execute her/his tasks

1. The participant enters the room, receives the consent form, and discusses it with the responsible researcher.

2. the researcher asks the participant whether she/he wants to participate to the data collection step.

3. If the participant decides to participate, he/she is being equipped with a GSR-sensor, a pulse belt and a temperature sensor.

4. The researcher starts collecting data, so to calibrate the sensors, for approximately ten minutes. During this time the participant is asked whether everything feels fine, and whether she wants to voluntarily answer questions about age, gender and weight or not.

5. When the calibration is done and if the participant has agreed to continue participating in the study, the researcher opens the SAM questionnaire on the app, and asks the participant to express how he/she is feeling by using the SAM app.

6. The researcher also checks, using her/his laptop, whether all the sensors are working properly and the data acquisition app is saving the measurements correctly in her/his Dropbox folder.

7. The researcher shows the baseline clip 2a to the participant.

8. The participant is asked to use the SAM once more to express how he/she is feeling, and if he/she has seen the film clip before.

9. The researcher instructs the participant to solve a cognitive distraction task for 5 minutes. The order of which cognitive task should be given now is the one according to the random order for the tasks generated before.

10. At the end of the cognitive task the participant is asked to express how he/she is feeling by using the SAM once again.

11. The researcher shows another emotional film excerpt following the random order for the films generated before.

12. The participant answers once again how he/she felt when watching the clip and if he/she has seen it before using the SAM.

13. The participant repeats tasks 9 - 12 until each cognitive task has been solved once, all the emotional films have been watched once, and each step has been followed by answering the SAM questionnaire.

14. If the participant desires, the responsible researcher discuss about exactly what we plan to do with their data/in the research.

15. The participant leaves the room.

A.5.3 Finalizing the data collection step by inserting some concluding information

1. The researcher opens the Apps/HealthDataCollector folder in his/her Dropbox and Creates a folder for the participant. He/she moves all the saved .txt recording files from the recent experiments to this folder.

2. The researcher fills out the ”information_about_participant.docx” with information about the participant and the experiment.

3. The researcher drags the .exe file into the different dataFA, dataFS, dataFT, dataFF, dataFB and plots the collected data to make sure that all the files is containing reasonable data by running it.

4. If the researcher hasn't done it already, he/she adds the link of his/her Apps/HealthDataCollector folder to the table in section A.7

A.6 List of all the links to the manuals and materials

Material	Link	Description
Manual for using the HealthDataCollector-app and sensors	link	The manual describes the technical aspect of the experiment. It is a step-by-step guide that explain how to use the app and the sensors, and how to handle the data that is being saved in Dropbox.
Cognitive distraction tasks	link	The folder which contains the materials for the four activities Maze solving, Sudoku, Geometric-shapes-copying and Find all a's
Consent form	link	All the participants need to sign a consent form before the experiment starts. Information about the study is also included in this document.
Folder with a form to fill out information about the participants and the experiment	link	The link directs you to a folder. All the measurements for the experiment will after an experiment be moved to a copy of this folder.
Film clip Tenderness	link	Life is beautiful, Tank scene
Film clip Amusement	link	When Harry met Sally, restaurant scene
Film clip Sadness	link	The Champ, ending scene
Film clip Fear	link	The Ring, Dream
Film clip baseline	link	Our planet - jungles, first 5 minutes

A.7 Links to the Dropbox folder that will contain all the measurements from the sensors

Researcher	Dropbox link
Rebekka	`link`
Damiano	`link`
Roya	`link`
Roxy	`link`
Steffi	`link`

Appendix B

Informed Consent

The following two pages present the consent form that we have been designing and using through our data collection campaign. The usage of this consent form has been approved by NSD.

Title of the study

DETECTION OF SPECIFIC EMOTIONS

- *based on physiological signals regulated by the autonomic nervous system.*

Information about the study

You are being asked to take part in a research study. Before you decide to participate in this study, it is important that you understand why the research is being done and what it will involve. Please read the following information carefully. Please ask if something is unclear or if you need more information.

You will by participating in this study contribute with health data that later will be analyzed and processed. The types of health data that is being collected is listed below.

- GSR: Galvanic skin response
- Pulse measurements: ECG-measurements, PPG-measurements, RR-intervals and heart rates.
- Skin temperature.
- Motion-based activity by using an accelerometer.

You will also be asked about your age, weight and gender. You have the right to not specify or, for age and weight, specify an interval rather than the specific number.

The purpose of the study is to explore the possibilities of automatic emotion recognition solely by processing the health data described above. You will be watching several movie clips that will trigger different emotions while your physiological signals are collected. *We will therefore ask you to reconsider your participation if you have any conditions or are in a state that makes you sensitive to some types of film scenes.*

The research is primarily conducted by the Norwegian University of Science and Technology(NTNU) in cooperation with Otto-von-Guericke University Magdeburg, Germany. The data collection process will partly take place in Magdeburg following the same protocol as at NTNU. The collected data will be shared across the universities for further joint research.

All collected data will be de-identified. The collected data CANNOT be traced back to the individual participant.

Do not hesitate contacting a psychologist if you experience any type of mental health problems.

Sit Psykososial helsetjeneste
Tlf: +47 73 53 86 30
(mon. - Fri.: 08.30-11.30, 12.30-15.30)

Adr: Bregnevegen 65 7050 Trondheim

Declaration of Consent

for participation in the research study:

DETECTION OF SPECIFIC EMOTIONS
- *based on physical signals regulated by the autonomic nervous system.*

- I have been informed of the study. I have read the written information. I have had the opportunity to ask questions about the study. I have been able to think about my participation in the study, which is completely voluntary. I have the right to withdraw my consent and leave the study at any time without having to give a reason.

- I am aware and agree that my personal data as specified in the written information, may be used for the study.

- I am aware and agree that my de-identified data can be used for further scientific research both in Norway and abroad.

- I agree to participate in the study.

Appendix C

Manual for using HDC

C.1 Introduction

This manual describes how to use the iOS application HDC and a set of sensors to collect pulse-, temperature and skin conductance data.

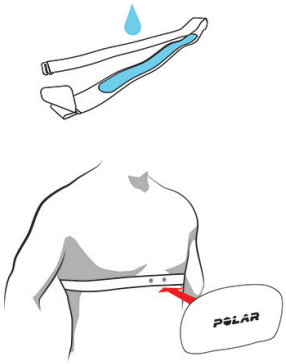
C.2 Sensor set

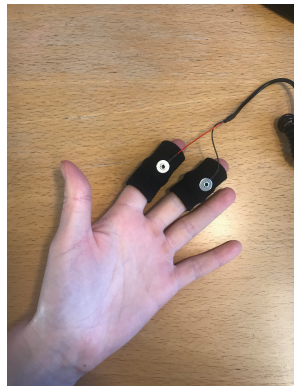
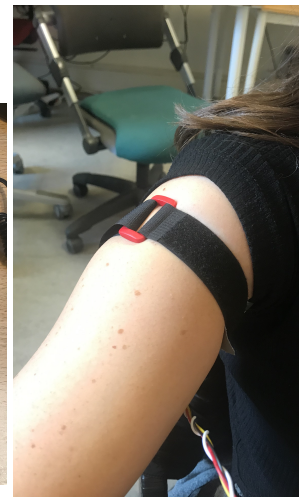
The set consist of three different sensor modules which are placed at different body parts.

- Polar pulse belt is fastened under the chest with the electrodes right in the middle of the chest as shown in the picture below. The electrodes are moistened with water. Make sure the fingers that you use for the GSR sensor are dry.

- The GSR sensor is attached to the middle finger and the index finger as shown in the picture below.

- The temperature sensor is fasten around the arm so that the sensor(black half-cylinder sticking up from the tape) is placed in the armpit. Make sure that the arm is held close to the body while recording and calibration.



C.3 Calibration

Both the temperature sensor and the GSR sensor need calibration before recoding. The longer time for calibration, the better. But it should calibrate at least 10 minutes. Attach the GSR- and the temperature sensor to the left arm if the participant is right handed and to the right if he/she is left handed.

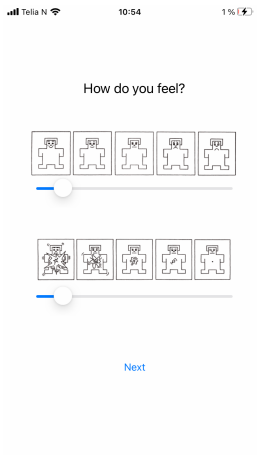
C.4 Dropbox

Before the recording starts, you need to make sure that the Dropbox-app is downloaded at you iphone and that you have enough free space.

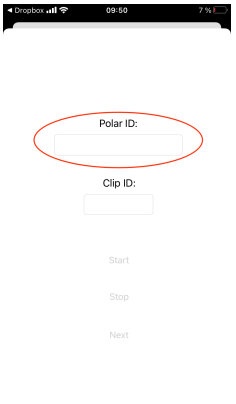
C.5 Before the recording

When all the sensors are attached to the body and calibrated, the recordings can start. The sequence of steps is the following:

1. Start the application.
2. Let the participant answer the questionnaire on how he/she feels. The upper 5 pictures represents the valence (positive/negative feeling). The lower 5 pictures represents the arousal (strength of the feeling)
3. Click the next button and follow the instructions of approving Dropbox's conditions.
4. Open the film clip that should be watched now or the cognitive task that should be executed now, but do not start it yet.



5. Write the ID of the polar belt in the upper input field. You will find the ID on top of the polar belt.



6. Write an identifier for the film clip or cognitive task in the lower input field. This identifier consists of just two capital letters, the first to indicate whether it is a film or a cognitive task, and the second to indicate the specific film clip or task. More precisely

   - FA for ”Film Amusement”
   - FS for ”Film Sadness”
   - FT for ”Film Tenderness”
   - FF for ”Film Fear”
   - FB for ”Film Baseline”
   - TA for ”Task find the A’s”
   - TG for ”Task Geometry”

- TS for ”Task Sudoku”
- TM for ”Task Maze”

7. Press the start button. Note that to make it possible that this button becomes blue (and thus pressable), you have to touch the white display after writing the clip number.

8. Be ready to start the clip after pushing the start button. Start the film clip immediately when the stop button is turning blue.

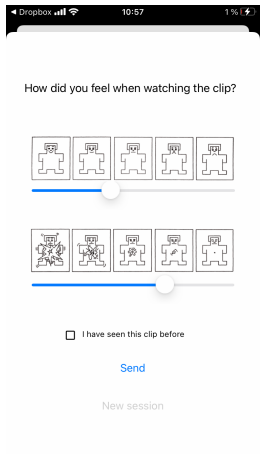
9. Watch the clip and make sure the volume is good. Use a set of earplugs.

10. Push the stop button when the clip is finished or the task executed

C.6 After the recording

1. Push the next-button

2. Move the sliders according to how you feel when watching the clip. Also click the checkbox if you have seen the clip before.

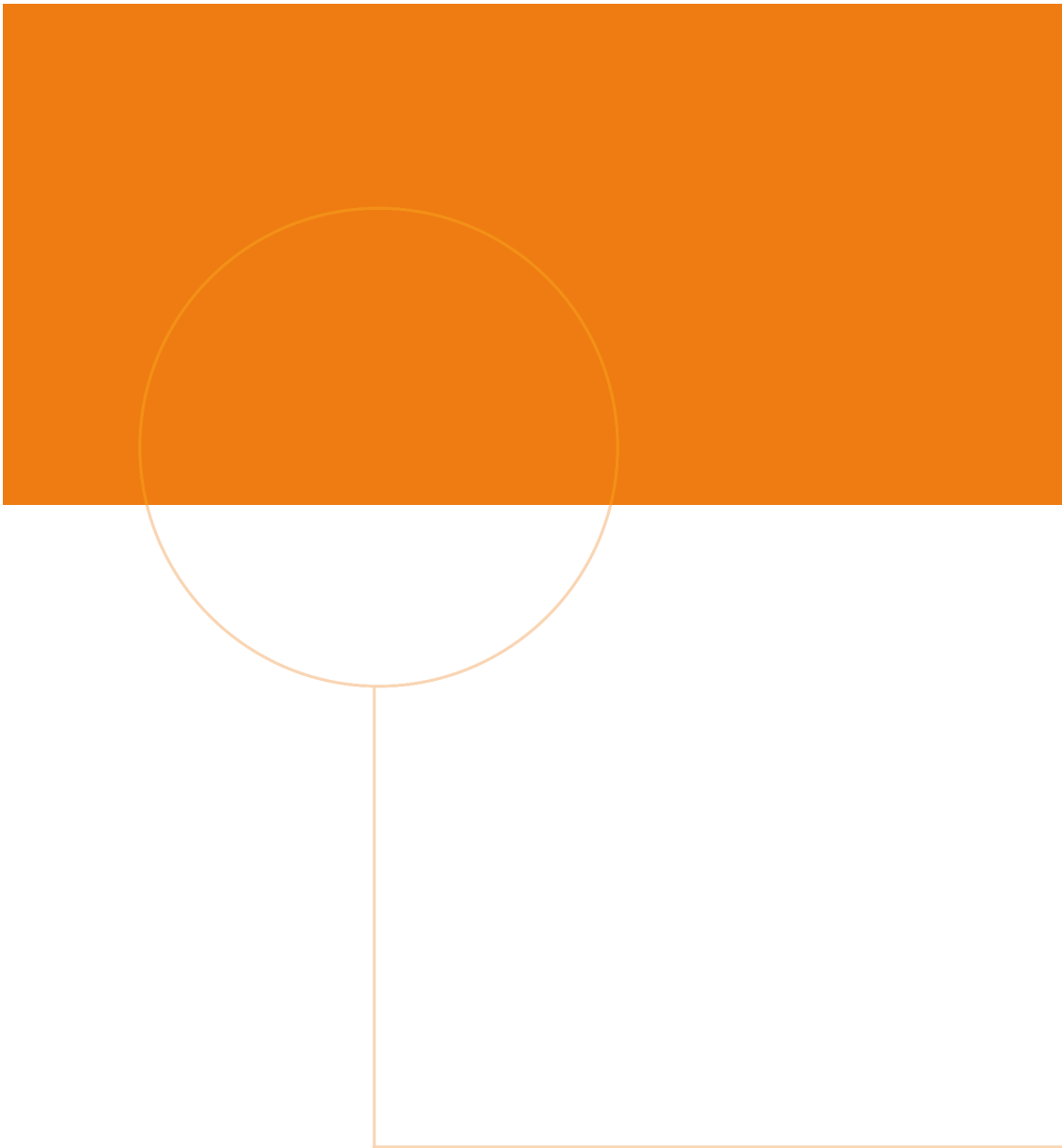


3. Push send. When the send-button is gray is all the sensor data saved in your Dropbox and you can quit the app or push the ”new session” button to start a new session.

4. When all the recordings for one particular participant is done, open the Apps/HealthDataCollector folder in your Dropbox. Create a folder for the participant and move all the saved .txt recording files to this folder. Include also the files found here: `link`

5. Fill out the ”information_about_participant.docx” with information about the participant.

6. Drag the .exe file into the different dataFA, dataFS, dataFT, dataFF, dataFB, dataTA, dataTG, dataTS, dataTM and plot the collected data to make sure that all the files is containing reasonable data by running it.



NTNU

Norwegian University of
Science and Technology