TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl für Steuerungs- und Regelungstechnik

Static and Dynamic Methods for Emotion Recognition from Physiological Signals

Robert E. W. Jenke

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr.-Ing. Eckehard Steinbach

Prüfer der Dissertation:

1. Prof. Dr.-Ing. Angelika Peer
   University of the West of England, Bristol, UK

2. Univ.-Prof. Dr.-Ing. habil. Gerhard Rigoll

Die Dissertation wurde am 05.05.2015 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 30.10.2015 angenommen.

Foreword

This thesis summarizes a substantial part of the work I did at the of Chair of Automatic Control Engineering (LSR) of the Technische Universität München. During this time, I had many exceptional opportunities to learn, experiment, and build new skills on both technical and personal levels. These experiences as well as the resulting thesis would not have been possible without the invaluable help and support of a number of people.

First of all, I would like to express my sincere gratitude to my advisor Prof. Angelika Peer for her sustained interest and support, for the high-quality discussions we had, for her detail-oriented advice, and for providing motivation and encouragement throughout the years, not least by allowing to assume responsibility and independent involvement in projects. I also like to thank Prof. Martin Buss for heading the LSR in a way that it offers an enriching and versatile environment with ample possibilities of personal development.

Most of my work has been embedded in the theme of the EU project *Virtual Embodiment and Robotic Re-Embodiment (VERE)*, which has been a wonderful international and interdisciplinary experience and has further augmented my work with several occasions to travel and visit other research labs. I have always enjoyed collaborations and discussions with the VERE partners, among whom are: Prof. Mel Slater (Universitat de Barcelona), who I highly esteem for coordinating the project, Prof. Verónica Orvalho, Xenxo Alvarez, and Bruno Oliveira (Universidade do Porto) as well as Prof. Tamar Flash and Irit Sella (Weizmann Institute), who I had the pleasure to collaborate with, and Bruno Herbelin (PhD), Dr. Michael Madary, and others for great discussions during the meetings. At this point, I also like to thank my colleagues on the VERE project: Mohammad Abu-alqumsan, Katrin Landsiedel, and Nikolas Martens for their fantastic cooperation. Another highlight during my PhD times has been a visit of the robotic summer school organized by the group of Prof. Schöner (Ruhr-Universität Bochum) and several insightful discussions that followed with Dr. Yulia Sandamirskaya.

The daily life at LSR was enjoyable not least due to many open-hearted colleagues. I thank Daniel Carton, Ken Friedl, Sheraz Kahn, Stefan Klare, Christian Landsiedel, Andreas Lawitzky, Nikos Mitsou, my roommate Roderick de Nijs, and Markus Rank for their friendship and help. I also had the pleasure of supervising more than 20 students, all of whom I thank for their contributions. Special thanks go to Marcel Debout, Florian Dörning, Lukasz Dobrogowski, Johannes Honold, Murad Lodhi, and Dongxue Wang for their efforts.

Last and most of all, I like to thank my partner Lena, my brother and my parents for their loving and endless support in every aspect of life.

Munich, April 2015   Robert Jenke

Abstract

Current efforts in human-machine-interaction aim at finding ways to make interaction more natural, e.g. by incorporating concepts of non-verbal communication. In this, knowledge about the user's emotional state is considered an important factor to correctly interpret actions and information. This research domain is termed "affective computing" and lies at the intersection of the fields of psychology and engineering. Methods of automatic and reliable estimation of affective states from various modalities has received much attention lately. In particular, emotion recognition from physiological signals is expedient, since it taps the pure, unaltered emotion in contrast to modalities like facial expressions which can be faked. It also does not require the user's attention, which is important if one is assessing emotions in interaction studies or parallel to other tasks.

In this thesis, static and dynamic methods for emotion recognition from physiological signals are investigated. In the first part, we take a closer look at static methods for emotion recognition in an effort to consolidate current approaches to feature extraction and selection in this field. Since the human body constitutes a complex interconnected system, the implications that arise in the presence of interacting features in feature selection problems are studied. We compare several state-of-the-art feature selection methods on an artificial dataset and evaluate their capabilities to detect different types of interactions. A systematic analysis of feature extraction and selection is produced and applied to EEG signals, for which a large quantity of features have been proposed without definitely clarifying which of these features and electrode locations are most suitable for emotion recognition. We review features from EEG signals and report the analysis of data from a user study with respect to this open question. The second part concerns itself with the development of a novel dynamic approach to emotion recognition, taking findings from modern emotion theory into account. In this, one of the key findings is that emotions are of dynamic nature, which has largely been overlooked in designing systems for emotion recognition. Therefore, a theory-incorporating dynamic framework for emotion recognition from physiological signals and other modalities is proposed. More specifically, we develop a gray-box model for emotion intensity estimation which allows the combination of theoretical knowledge with data-driven experimental approaches. A proof-of-concept user study demonstrates significant improvements of the proposed model to common methods and baselines.

Zusammenfassung

Gegenwärtige Bestrebungen in der Mensch-Maschine-Interaktion zielen darauf ab, die Interaktion, z.B. durch die Integration verschiedener Konzepte der nicht-verbalen Kommunikation, natürlicher zu gestalten. Dabei wird die Einsicht in den emotionalen Zustand des Benutzers als wichtig erachtet, um Aktionen und Informationen korrekt zu interpretieren. Der sich damit beschäftigende Forschungsbereich wird unter dem Begriff des "Affective Computing" zusammengefasst und kann als eine Schnittmenge der Bereiche Psychologie und Ingenieurwissenschaften formuliert werden. Methoden für die automatische und zuverlässige Erkennung des emotionalen Zustands aus verschiedenen Modalitäten haben zuletzt vermehrt Beachtung gefunden. Im Besonderen ist die Emotionserkennung von physiologischen Signalen vorteilhaft, da man, im Gegensatz zu andere Modalitäten wie Gesichtsausdrücken, welche verfälscht werden können, auf diesem Weg Zugriff auf die echte, unveränderte Emotion hat. Des Weiteren erfordert diese Modalität nicht die Aufmerksamkeit des Benutzers, was besonders in Studien, in denen Emotionen während Interaktionen oder parallel zu anderen Aufgaben ermittelt werden sollen, von Vorteil sein kann.

In der vorliegenden Dissertation werden sowohl statische als auch dynamische Methoden für die Emotionserkennung aus physiologischen Signalen untersucht. Im ersten Teil der Arbeit werden statische Methoden der Emotionserkennung betrachtet im Bestreben, die aktuellen Ansätze zur Merkmalsextraktion und -selektion zu verdichten. Nachdem der menschliche Körper ein komplexes, in sich vernetztes System ist, werden zunächst Implikationen, die mit dem Auftreten interagierender Merkmale im Zusammenhang der Merkmalsselektion verbunden sind, untersucht. Ein Vergleich von mehreren fachüblichen Methoden der Merkmalsselektion anhand eines künstlich erstellten Datensatzes beurteilt deren Fähigkeit, verschiedene Typen interagierender Merkmale zu erkennen. Ein systematischer Ansatz zur Analyse von Merkmalsextraktion und -selektion wird vorgetragen und auf EEG Signale angewendet, für die eine große Anzahl an möglichen Merkmalen vorgeschlagen wurden. Jedoch ist nicht abschließend geklärt, welche der Merkmale und welche Elektrodenpositionen am geeignetsten für die Emotionserkennung sind. Hinsichtlich dieser noch offenen Frage, wird ein Überblick über die Merkmale aus EEG Signalen sowie die Ergebnisse aus der Datenanalyse einer Benutzerstudie präsentiert. Der zweite Teil der Arbeit beschäftigt sich mit der Entwicklung eines neuartigen, dynamischen Ansatzes zur Emotionserkennung, welcher Erkenntnisse aus der modernen Emotionstheorie miteinbezieht. Ein zentrales Ergebnis ist die Feststellung, dass Emotionen dynamischer Natur sind, was beim Design von Systemen zur Emotionserkennung bislang jedoch weitgehend übersehen wurde. Daher wird eine theoretisch motivierte Architektur zur Emotionserkennung von physiologischen Signalen (wie auch von anderen Modalitäten) vorgestellt. Im Speziellen wird ein sogenanntes "gray-box" Modell zur Schätzung von Emotionsintensitäten entwickelt, welches die Kombination von theoretischem Wissen mit datengetriebenen, experimentellen Ansätzen erlaubt. Die Anwendung in einer Benutzerstudie demonstriert eine signifikante Verbesserung der Ergebnisse aus dem vorgeschlagenen Modell im Vergleich zu Standardmethoden.

Contents

Notations

Abbreviations

NVC	non-verbal communication
HMI	human-machine interaction
aBCI	affective brain-computer interface
DAQ	data acquisition
EEG	electroencephalography
ECG	electrocardiography
fMRI	functional magnetic resonance imaging
GSR	galvanic skin response
ANS	autonomous nervous system
SNS	somatic nervous system
CPM	component process model
SEC	stimulus evaluation check
IAPS	international affective picture system
SAM	self-assessment-manikin
VAD	valence-arousal-dominance
DFT	dynamic field theory
DNF	dynamic neural field
SVM	support vector machine
LDA	linear discriminant analysis
i.i.d.	independently and identically distributed
FS	feature selection
ES	effect size
mRMR	minimal-redundancy-maximal-relevance
SFS	sequential forward selection
ERP	event related potentials
NSI	non-stationary index
HOC	higher order crossings
DWT	discrete wavelet transform
FFT	fast Fourier transform
STFT	short-time Fourier transform
PSD	power spectral density
HOS	higher order spectra
HHS	Hilbert-Huang spectrum
EMD	empirical mode decomposition
MSCE	magnitude squared coherence estimate

Conventions

Scalars, vectors, and matrices
Scalars are denoted by lower or upper case letters. *Vectors* are denoted by bold lower case letters, as the vector $\mathbf{x}$ is composed of elements x_i. *Matrices* are denoted by bold upper case letters, as the matrix $\mathbf{X}$ is composed of elements x_{ij} (i^{th} row, j^{th} column).

x or X	scalar
$\mathbf{x}$	vector
$\mathbf{X}$	matrix
$\mathbf{X}^T$	transposed of $\mathbf{X}$
$\mathbf{X}^{-1}$	inverse of $\mathbf{X}$
$f(\cdot)$	scalar function
$\hat{x}$	estimated or predicted value of x
$\dot{x}, \ddot{x}, \ldots$	first, second, $\ldots$ time derivative of x

Subscripts and superscripts

$(\cdot)^*$	complex conjugate
$(\cdot)_\xi$	associated with time signal ξ
$(\cdot)_\mathbf{x}$	associated with feature $\mathbf{x}$
$(\cdot)_{u/i/m/v}$	associated with layer $u/i/m/v$, respectively
$(\cdot)^c$	associated with samples belonging to class c

Symbols

General

A_{bdw}	bandwidth accuracy
b	bandwidth of A_{bdw}
C	number of classes
c	class label
d	Cohen's effect size
$\mathbf{E}$	within sum of square and cross product (SSCP) matrix
F	number of features
f	frequency
f_S	sample frequency
$\mathbf{H}$	between sum of square and cross product (SSCP) matrix
I	emotion intensity
λ	eigenvalue
MAE	mean absolute error
MSE	mean squared error

$\boldsymbol{\mu}$	mean vector
μ	mean
N	number of samples per class
N_T	total number of samples
n	sample
$\boldsymbol{\nu}$	eigenvector
r	correlation coefficient
$\mathscr{S}$	subset of selected features
$\boldsymbol{\Sigma}$	covariance matrix
σ	standard deviation
T	number of time samples
t	time
ϑ	emotion quality
$\mathbf{X}$	feature matrix (all samples of all features)
$\mathbf{x}_i$	feature vector (all samples of i-th feature)
x	one sample of a feature
$\mathscr{X}$	set of all features
$\Xi(f)$	Fourier transform of $\xi(t)$
$\xi(t)$	time signal
$\mathbf{y}$	target vector (class information of all samples)
y	class information of one sample
z	number of features in $\mathscr{S}$

Chapter 2-specific

e	discrete emotion
d_a	arousal dimension coordinate
d_v	valence dimension coordinate

Chapter 3-specific

a	scaling factor
D	maximal-relevance criterion
f^2	univariate effect size measure
$I(\mathbf{x}, \mathbf{y})$	mutual information between $\mathbf{x}$ and $\mathbf{y}$
k	number of nearest hits/misses
Λ	Wilk's statistic
m	number of selected samples
$p(\mathbf{x})$	probability density function of $\mathbf{x}$
q	adjustment variable
R	minimum-redundancy criterion
s	step size
σ_m	between class standard deviation
θ	Roy's statistic

U	Pillai's trace
V	Hotelling criterion
W_i	quality weight of i-th feature
$\mathrm{x}_{H/M}$	nearest hit/miss, respectively

Chapter 4-specific

A_ξ	activity
$A_i(t)$	amplitude of the i-th IMF
$\mathrm{Bic}(\cdot)$	bicoherence
$\mathrm{Bis}(\cdot)$	bispectrum
C_ξ	complexity
C_{ij}	MSCE between ξ_i and ξ_j
D_{ij}	j-th detail coefficient at ith decomposition level
δ_ξ	1st difference
$\bar{\delta}_\xi$	normalized 1st difference
E_{band}	energy of a frequency band
$e_K(t)$	residue of HHS
FD_ξ	fractal dimension
g	time interval
γ_ξ	2nd difference
$\bar{\gamma}_\xi$	normalized 2nd difference
H	number of zero crossings
$\mathrm{IMF}_i(t)$	i-th intrinsic mode function
$\mathfrak{I}_k$	high-pass filter of order k
J_i	number of detail coefficients at ith decomposition level
K	number of IMFs
$L_m(g)$	m-th set computed for the Higuchi algorithm
l	number of decomposition levels
M_ξ	mobility
m	initial time
∇	difference operator
P_ξ	power
$P(f)$	power spectrum
$P_{ij}(f)$	cross power spectral density
REE	recursive energy efficiency
RMS	root mean square
$\theta_i(t)$	phase of the i-th IMF
$Z(t)$	zero-mean time signal
$\delta, \theta, \alpha, \beta, \gamma$	frequency bands of EEG signals

Chapter 5 & 6-specific

$\tilde{A}$	activity of predicted SEC effect
β	slope of sigmoid function
c	coupling weigths
ϵ	model output threshold
h	resting level
$I(\hat{y}, y)$	$0 - 1$ loss function
i	intensity layer
m	memory layer
o_i	model value of the i-th sample
$\omega(\cdot)$	interaction kernel function
$\mathbf{p}$	parameter vector
R^2	evaluation statistic for linear regression
S	input to DNF
σ_u	uncertainty measure
τ	time constant
u	input layer
v	coupled layer
$W(x, y)$	entries of a weighting matrix
x, y	location of DNFs

List of Figures

List of Tables

1 Introduction

1.1 Motivation

The development of intelligent machines has be a major goal for decades, but machines were long believed to be of purely logical nature. In every-day interactions among human beings (Human-Human-Interaction), however, a large part of information about our own state and about the state of the interaction is shared using non-verbal communication (NVC) [1]. This NVC, which can come in forms of gestures or tone of voice, for example, is important for the correct interpretation of actions as well as communication. Consequently, when humans interact with machines and robots (Human-Machine-Interaction, HMI), they would like to apply a similar social model [2]. When presumably intelligent machines do not react to NVC or do not utilize NVC correctly, interaction may feel uneasy and may lead to discomfort [3].

Thus, current efforts in HMI research aim at making interaction more natural by trying to meet the implicit expectations held by human users, which in many cases are based on experience with other humans [4]. In this, knowledge about the emotional state of a user is considered an important factor, since emotions affect our cognitive processes and thus, our interactions and NVC, on many levels [5]. Although the reliable recognition of the current emotional state is prerequisite for successful application, the topic of emotion recognition has largely been overlooked for a long time.

The improvement to HMI when including emotions are conceivable on several levels [6]. On the one hand, machines that understand the user's emotion can react in a more appropriate manner, e.g. by varying the parameters or interface options accordingly. For example, in the currently flourishing e-learning system, the level of difficulty can be adjusted when detecting frustration of the user. Another related example is game design, in which emotional states are used to provide hints and clarifications or are similarly used to adjust parameters for difficulty [7]. On the other hand, HMI will be more effective when machines include the concepts of emotion in their own expressions, i.e. consider usage of emotion as an additional modality of interaction by adjusting actions like gestures, facial expressions or the tone of speech themselves. An e-learning system, for example, could incorporate jokes to lighten the mood in order to make the interaction more fun and engaging. It is also imagined to train desired emotional states like relaxation by changing the environment according to the detected emotional states, e.g. Brainball [8] and later developments [9].

Another aspect which motivates research in emotion recognition is that of gaining a better understanding of emotions themselves. In the course of building and developing systems that can correctly identify a large variety of emotional states, a strong connection of the fields of engineering and psychology is natural. In this process of study, the understanding of the underlying mechanisms of emotions inevitably increases for the best of both fields.

Since Picard coined the term "affective computing" in 1995 [10], methods of automatic and reliable estimation of affective states from various modalities have received much atten-

tion. Modalities can be grouped into facial expressions, body gestures and posture, speech, and physiological signals. Each modality carries different information, which can complement each other when combined (multi-modal emotion recognition). In particular, emotion recognition from physiological signals is expedient, since it taps the pure, unaltered emotion in contrast to modalities like facial expressions which can be faked. It also does not require the user's attention, which is important if one is assessing emotions in interaction studies or parallel to other tasks.

Physiological signals cover both brain signals and peripheral physiological signals of the autonomous nervous system (ANS), which include heart activity, respiration, temperature, skin conductance, and others. The recent interest in applications using brain signals has coined the term Brain-Computer-Interface (BCI) and the specialized field of affective BCI (aBCI) [7]. The latter is rated as a passive BCI, i.e. it enriches the interaction without the need for voluntary effort of the user [11]. Work that combines both brain signals and peripheral signals is also referred to as a Brain-Body-Computer-Interface (BBCI) [204].

1.2 Problem statements and challenges

The ultimate goals of research in emotion recognition are sophisticated applications in HMI as listed above. To reach the level of perfection in this field required to achieve these goals, a number of major challenges have yet to be overcome. Particularly in this interdisciplinary field of study, it is important to understand current challenges from both an engineering and psychological point of view. Only then can the important issues be addressed in systematic ways which lead to meaningful advancements of the field. The challenges underlying the methods developed and investigated in this thesis are summarized in the following sections.

Feature selection methods and interacting features

Since the human body is a complex interacting system, the physiological signals that are tapped to recognize emotions are likely to be interacting as well. When a large number of features extracted from these signals are available, feature selection (FS) techniques are applied to reduce this space without loosing meaningful information. A quantity of FS methods has been developed in machine learning, but research on the implications of interacting features is limited. Particularly, the question of which FS method can detect which types of interaction under fully controllable conditions has not received sufficient attention.

Features for emotion recognition

Emotion recognition from physiological signals allows the direct assessment of the "inner" state of a user. Especially for brain signals, many methods for feature extraction have been studied and the selection of both appropriate features and electrode locations is usually based on neuro-scientific findings. Their suitability for emotion recognition, however, has been tested using a small amount of distinct feature sets and on different, usually small datasets. A major limitation is that no systematic comparison of features exists.

Modeling emotion dynamics

To date, emotion recognition has mostly been carried out using static machine learning methods. However, recent developments of emotion theory pointed out the dynamic nature of emotions. While emotion dynamics have been studied on various levels, current approaches to emotion recognition from physiological signals do not address the fact that emotions are dynamic processes. Thus, it is of special interest to develop models for emotion recognition that can incorporate these findings.

Experimental designs

One barrier related to the development of dynamic models for emotion recognition is the increased difficulty to design and carry out pertinent user studies. Especially the dynamic measurement of emotion on a time continuous scale is still not definitely solved. Moreover, the design of experiments to induce fine grained differences, for example, of the intensity of emotions is challenging.

Foundations of emotion recognition (ER)	
Emotion models, emotion induction and labeling, common processing pipeline	Ch. 2

Part I: Static ER	**Part II: Dynamic ER**
Feature selection (FS) • Discussion of interacting features • Comparison of FS methods Ch. 3	**Dynamic model** • Incorporation of appraisal • Dynamic gray-box model Ch. 5
Emotion recognition from EEG • Review of feature extraction • Analysis on real-world dataset Ch. 4	**Intensity estimation from GSR** • Novel experimental design • Proof-of-concept study Ch. 6

Figure 1.1: Outline of the thesis.

1.3 Contributions and thesis outline

The thesis is structured as outlined in Fig. 1.1. Chapter 2 introduces fundamental concepts of emotion recognition from a psychological, experimental, and engineering viewpoint. The main content is presented in chapters 3 through 6 and is divided into two parts. In the first part of the thesis we take a closer look at static methods for emotion recognition in an effort to consolidate current approaches to feature extraction and selection in this field. The second part concerns itself with the development of a new dynamic approach to emotion recognition, taking novel findings from emotion theory into account. With respect to the problem statements provided above, the contributions made within this thesis are summarized in the following sections.

A comparison of feature selection methods detecting interacting features

In Chapter 3, we consider the implications that arise in the presence of interacting features in feature selection problems. We deepen the understanding of this topic by first giving an overview of FS methods and present four different types of feature interaction in multi-class problems. Further, we study an academic example with the aim to compare selected FS methods w.r.t. their capabilities of detecting and exploiting feature interactions. We find that the FS methods differ in their ability to detect different types of interaction, which should encourage the use of multiple FS methods at the same time to increase feature stability.

Feature extraction and selection from EEG

In Chapter 4, we address the problem of FS from physiological signals in a systematic way applied to the example of electroencephalography (EEG) features. Therefore, we review feature extraction methods for emotion recognition from EEG based on over 30 studies. An experiment is conducted comparing these features using machine learning techniques for feature selection on a self recorded dataset. Results are presented with respect to performance of different feature selection methods, usage of selected feature types, and selection of electrode locations. Features selected by multivariate methods slightly outperform univariate methods. Advanced feature extraction techniques are found to have advantages over commonly used spectral power bands. Results also suggest preference to locations over parietal and centro-parietal lobes.

A dynamic model incorporating appraisal into emotion recognition

In Chapter 5, the topic of dynamics of emotions is addressed. In this, we suggest a theory-incorporating framework for emotion recognition from physiological signals and other modalities. The work concerns itself with the development of a gray-box model for dynamic emotion intensity estimation that can incorporate findings from appraisal models, specifically Scherer's Component Process Model. It is based on the Dynamic Field Theory which allows the combination of theoretical knowledge with data-driven experimental approaches.

Emotion intensity estimation from physiological signals

In Chapter 6, we describe a user study that was carried out as an example to verify the model proposed in Chapter 5. In this, we provide the experimental design as well as a comparison of evaluation measures for continuous emotion recognition. The proposed model is applied to estimate the intensity of negative emotions from physiological signals. Results show significant improvements of the proposed model to common methodology and baselines. We suggest multiple extensions to the flexible framework which opens a wide field of experiments and directions to deepen the understanding of emotion processes as a whole, in an effort to advance emotion recognition from physiological signals.

2 Fundamentals of emotion recognition

Summary. *Research in emotion recognition is highly interdisciplinary work. Concepts of both psychological principles of emotion and computational methods for emotion recognition are needed to make meaningful advances in this field. To provide the reader with a short introduction to the fundamentals of this field, among the topics presented in this chapter are*

- *definitions and models of emotion and affect,*

- *methods for emotion induction used in user-studies,*

- *the machine learning process commonly applied in emotion recognition.*

The field of emotion recognition is interdisciplinary in its nature and requires both knowledge of the psychological principles of emotion and expertise in machine learning. In this chapter, we introduce the fundamentals of emotion recognition including concepts from each discipline and discuss how they are applied in this thesis. While we focus on emotion recognition from physiological signals, obviously most of the fundamental concepts apply to other modalities like speech, or gestures as well. A review of emotion models along with definitions of emotion and affect is presented in Section 2.1. We introduce methods of emotion induction needed to carry out experiments for data collection in Section 2.2. The general processing pipeline of emotion recognition applying machine learning methods is briefly summarized in Section 2.3.

2.1 Emotion models

Emotions have been investigated in a variety of ways. Besides the multitude of every-day interpretations of emotion, several scientific comprehensions have been proposed. Among those, for example, Frijda described the core of an emotion as the change in action readiness [12]. Another notion is that emotions are states elicited by rewards (you work for to get) and punishers (you work for to escape) [13]. Commonly mentioned is the working definition derived by Kleinginna and Kleinginna from a collection of definition used in literature [14]:

> Emotion is a complex set of interactions among subjective and objective factors, mediated by neural/hormonal systems, which can (a) give rise to affective experiences such as feelings of arousal, pleasure/displeasure; (b) generate cognitive processes such as emotionally relevant perceptual effects, appraisals, labeling processes; (c) activate widespread physiological adjustments to the arousing conditions; and (d) lead to behavior that is often, but not always, expressive, goal

directed, and adaptive.

But there is no universal definition of emotion. As Smith and Lazarus quoted from the Dictionary of Psychology, emotions are "differently described and explained by different psychologists, but all agree that it is a complex state of the organism", usually including strong feeling and action tendencies [15]. Rather, several emotion models exist and are still argued over, which define emotion in a context-specific way.

Early on, the definitions of terms related to mental states like *emotion*, *mood*, *affect* and *feeling* have been subject of debate [16]. In today's theoretical work, the term *emotion* usually refers to the whole emotion process including all components of the organism involved [17]. An affective state can be regarded as the subjective experience of an emotion, i.e. the *feeling* itself which is not necessarily directed at an object or event [18], [19]. Further, moods are generally considered to last much longer, i.e. hours or days, than emotions, which commonly last for seconds or minutes [17], [20].

Emotion models are commonly divided into three major types: discrete or categorial models, dimensional models, and appraisal models. While the first two represent the affective state of a person, the latter aims at describing the emotion process as a whole. Each type of model is introduced below together with prevalent examples and their main representatives.

2.1.1 Categorial models

Categorical or discrete emotion models are closest to what we use in our everyday lives when we use a single word to describe an affective state. Paul Ekman is probably the most popular representative of this model since Darwin. He derived the concept of six basic emotions *happiness*, *anger*, *disgust*, *sadness*, *anxiety*, and *surprise* from universally recognized facial expressions [21], [22]. Over the years, several expressions were added and removed from the list. For example, he presents *contempt* as a seventh pan-cultural facial expression [23]. This inconsistency is also a source of criticism of these models, since it is not agreed upon which emotions are basic and which are not [24]. Other major models of basic emotions have been reviewed by Tracy and Randles in 2011 [25].

In emotion recognition using classification methods, discrete emotion models account for the largest parts of studies carried out. From a computational point of view, these models are easy to implement, but it is difficult to model relations between the discrete states.

2.1.2 Dimensional models

Dimensional models have been introduced, which overcome the limitation of relating affective states to each other by providing a distance between them, i.e. they allow for computationally interpretable relations between emotional states. In this, the VAD model suggested by Russell and Mehrabian is prevalent [26], which spans a three-dimensional emotion space of independent and bipolar dimensions *pleasure*-displeasure (later termed *valence*), degree of *arousal*, and *dominance*-submissiveness (short: VAD space). A 3D numerical vector denotes the location of an emotion within this space, so that discrete emotions have corresponding VAD coordinates. This model has gained attention in the field of emotion recognition lately, both in studies applying regression methods and studies that discretize dimensions into few
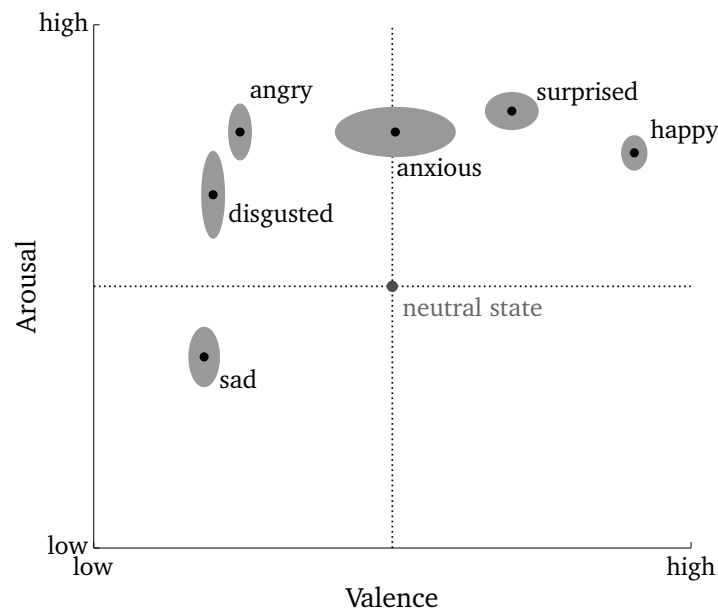
Figure 2.1: Ekman's basic emotions and their coordinates in valence-arousal space charac-
terized by the mean (solid point) and variance (shaded area).

areas, so that standard classification methods can be applied. Often, the third dimension
is omitted, resulting in the VA space. This circumplex model, which was discussed by Rus-
sell in 1980, uses the dimension *valence* and *arousal* and finds a circular arrangement of
affect words which are described by angles [27]. This representation of polar coordinates
has found less attention in emotion recognition, though. An advantage of including the *dom-
inance* dimension is that some emotions that are overlapping in VA space, e.g. fear and anger,
become distinguishable in VAD space. When considering subtle emotional differences, more
importance may be given to the three-factor version of this model in future research, as also
argued by Broekens et al. [28].

Several other dimensional models like Plutchik's emotional wheel [29] exist. Yik et al. re-
cently integrated many of these into a 12-point circumplex structure of core affect [30].

In this thesis, we adopt the view of the VAD model and use the mapping of discrete emo-
tional states to defined areas in VAD space as given by Russell and Mehrabian [26] to dis-
cretize the space. To illustrate this relation, Fig. 2.1 depicts the coordinates (mean and vari-
ance) of Ekman's basic emotions in VA space. We draw upon this mapping when inducing
emotions using pictures, which is further discussed in Section 2.2, and for classification of
physiological patterns (see Chapter 4).

2.1.3 Appraisal models

Appraisal models take on a more general view: They define emotions as processes, derived
from the notion that emotional episodes involve "changes in a number of organismic sub-
systems or components" [31]. In addition to merely providing a representation of emotion,
these models can explain underlying cognitive mechanisms of the emotional process in hu-
man beings by including appropriate components. All postulate an individual's subjective

Event

Appraisal

SNS

ANS

Motivation

Feeling
Quality
Intensity
Duration
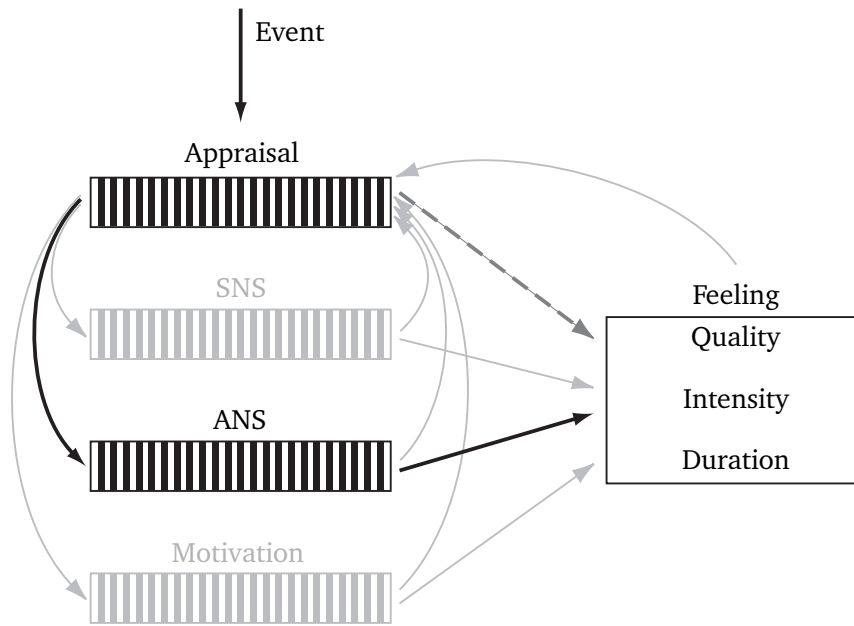
Figure 2.2: Mechanism of component integration (redrawn based on [32])

evaluation, i.e. appraisal, of the significance of an event [32]. They differ in the characteristic and number of components and appraisal dimensions considered. Leading appraisal theorists are Lazarus, Frijda, OCC, and Scherer [33]. The theory of the latter introduced in the next section is particularly interesting for application to emotion recognition [34], as we will show in Chapter 5.

The Component Process Model (CPM)

Scherer proposes a recursive model of emotions, called the Component Process Model (CPM), in which emotions are emergent processes that constitute themselves from the appraisal of an event in interaction with different components [32]. Five functional components are involved (see also Fig. 2.2): the cognitive component which includes the appraisal processes, the motivational component, the physiological efferent effects in the automatic nervous system (ANS), the motor expression component in the somatic nervous system (SNS), and the subjective feeling component. The cognitive component comprises a sequence of four stimulus evaluation checks (SECs): relevance (novel/not-novel, pleasant/unpleasant, goal relevant/not-relevant), implications (conductive/obstructive, urgent/not-urgent), coping (control/no-control, high power/low power, adjustment possible/adjustment not possible), and normative significance (standards (internal/external) surpassed/standards violated). Each influences the response of the other components. Subjective feeling takes on a special role, since it represents the conscious experience of an emotion, i.e. the affective state. It is composed of emotion quality, intensity and duration, which are determined from the current state of each of the other components. More details of the model are discussed, for example, in [32].

Although the CPM also includes predictions for facial expressions, voice, and body move-

ments, e.g. certain (facial) action units are modulated with changes in one of the evaluation checks, this model has found little attention in emotion recognition so far. In Part II of this thesis, we produce an approach to incorporate appraisal into emotion recognition, with the aim of enhancing recognition with the understanding of the underlying process, that this model offers.

2.1.4 Unifying model for representation of affect

Clearly, the various existing emotion models introduced above differ in their conceptual characteristics. Yet, each contains a representation of the affective state. In search for a unified computational representation, we can consider the following relations between the models.

Both discrete and dimensional models represent only the affective state rather than the whole emotion process. In the CPM, the affective state is modeled by the subjective feeling component. The interpretation of affective state as emotion quality and intensity, as given in the CPM, is not far from the aforementioned dimensional models: When using polar coordinates instead of Cartesian coordinate (d_v, d_a) for an affective state, the angle ϑ (ranging from $[0; 2\pi]$) denotes the emotion quality and the length of the vector I represents the intensity of the emotion (see Fig. 2.3). Neutral state is located in the center of the dimensional space, such that the duration of an emotion episode is determined by the time of non-zero intensity values. In the valence-arousal space, intensity is sometimes mistaken (at least in our idea and that of Reisenzein [35]) and equated with the arousal. Reisenzein showed, however, that the length of the vector is the only tenable option. A similar relation of intensity representation in a two dimensional plane has been proposed by Scherer in the Geneva Emotion Wheel [17].

As already mentioned above, a discrete emotion label e can also be represented in dimensional space by applying existing mappings [26].

2.2 Emotion induction

Producing reliable data for research needs special attention in emotion recognition, since emotions comprise spontaneous and short episodes, which occurrences in real life are usually not well predictable. This makes it particularly difficult to collect recordings of physiological or other signals in emotional situation. Therefore, emotion elicitation has been studied with the aim to provide tools and methods to create reliable but economically affordable datasets under experimentally controlled conditions. The existing approaches can be summarized in ascending order of their degree of control over the experiment:

- Induced behavior: Natural real-life reactions can be achieved by constructing very realistic situations using cover stories or secondary tasks like manipulated games [36]. However, while this method requires a lot of planning the reactions may vary strongly. Furthermore, ethical reservations remain, for example, when deceiving subjects.

- Free recall: Reactions similar to real-life situations are expected to arise by asking subjects to imagine or relive emotional situations or create emotional mental images, e.g. writing short assays about an event in the past [37].
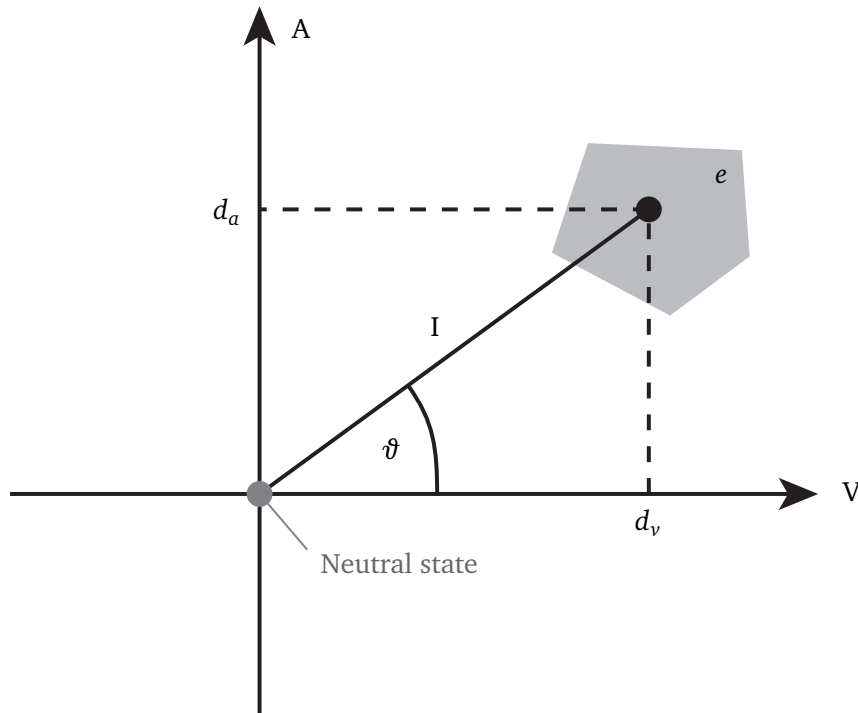
Figure 2.3: A unified computational representation of the affective state for discrete emotion models, dimensional emotion models, and CPM's subjective feeling.

- Guided imagination: Similarly, elicitation can be achieved by telling an imagination narrative or situational story in which the subject is instructed to picture herself in [38]. A disadvantage of all methods mentioned so far is the lack of specification of the onset of an emotion.

- Auditory stimuli: Direct presentation of emotional stimuli overcome this drawback by providing a defined onset. Sounds that have emotional impact are one form of emotional stimuli which are presented to subjects for emotion elicitation. A well-known standardized database, the International Affective Digitized Sounds (IADS), has been developed by Bradley and Lang [39]. Several studies have also used music to elicit emotion, but it is generally questioned whether music can reliably induce emotion [40].

- Visual stimuli: The majority of studies uses visual stimuli to induce emotion in a controlled way. In this, the International Affective Picture System (IAPS) is most popular due to both its size and its well-studied affective ratings of each picture [41]. Additionally, movie clips have been proposed, which represent very complex stimuli. For example, Koelstra et al. selected music videos to induce emotion [42]. Baveye et al. recently provided a database of labeled movie clips, named LIRIS-ACCEDE [43].

- Posed behavior: When the environment does not allow for successful emotion elicitation, it is possible to pose emotional reactions for some modalities like facial expressions or body posture and motions, as considered by Karg et al. [208], for example. This method is difficult to apply for physiological signals, and hence, the validity of the

recordings in other modalities can even be questioned, since it does not provoke an actual emotion but only action tendencies deliberately expressed by the subject.

- Administration of drugs: Last, and only mentioned for the sake of completeness, the application of drugs could possibly be used to trigger bodily emotional responses with high control over the experimental condition. But without a doubt, this method raises strong ethical concerns.

Of the various methods introduced above, we selected IAPS pictures to induce emotions, since it provides a good balance between control over the experiment and reliable emotion induction. Further, the extensively studied labels of the pictures facilitate an a-priori selection of stimuli in the experimental design phase. Although reactions may not be as intense as in some real-life situation, our experience with this method has been encouraging. Besides the selection of an appropriate induction method, its integration into a proper experimental design and appropriate instructions to the subjects is paramount to carrying out successful experiments. The experimental protocols put in place for experiments within the scope of this thesis are explained in detail later on (see Section 4.4 and Section 6.1).

2.2.1 Labeling of physiological data

In modalities for emotion recognition like facial expression, which focus on the expressive (re-)actions connected to an emotion, labels can be generated, for example, by a specialist or multiple viewers, each rating the expression w.r.t. its emotional content. When dealing with physiological data from affective experiments to recognize emotions, the recorded signals do not constitute an expression, but rather internal changes of the organism as part of an emotional experience. In this case, two different labeling methods can be applied [44]:

1. Operator labeled: A common way of labeling is to assign a discrete label (e.g. *happy*) to a pre-defined segment of the recording in accordance with the targeted induced emotion. While this method is very simple and avoids issues arising from the subject's reporting ability, the assigned label might be significantly different to the actually experienced affective state. Consequently, if operator labels are desired, a control via a self-report is necessary in order to validate the induction success, e.g. by using the Self-Assessment-Manikin test (SAM) [45].

2. User labeled: Since each human might react slightly different to stimuli (individual bias of emotion interpretations), it can be argued that the experience should be used as a label [46]. The advantage given a good reporting ability of the subject, is that it is closer the subjective feeling state. On the downside, problems of individual scale, difficulty of representation, subjects' reporting ability, and issues of time-continuous measurement (which we discuss in Section 5.2.4) arise.

In the first part of the thesis (Part I), we use targeted induced emotions as class labels in combination with a SAM test to validate induction success, i.e. sufficiently high agreement between targeted and self-reported affective state had to be achieved. In the second part of the thesis (Part II), where we focuses on dimensions of the subjective feeling on a continuous scale, we had subjects report their experienced affective states online, which are subsequently used as labels.
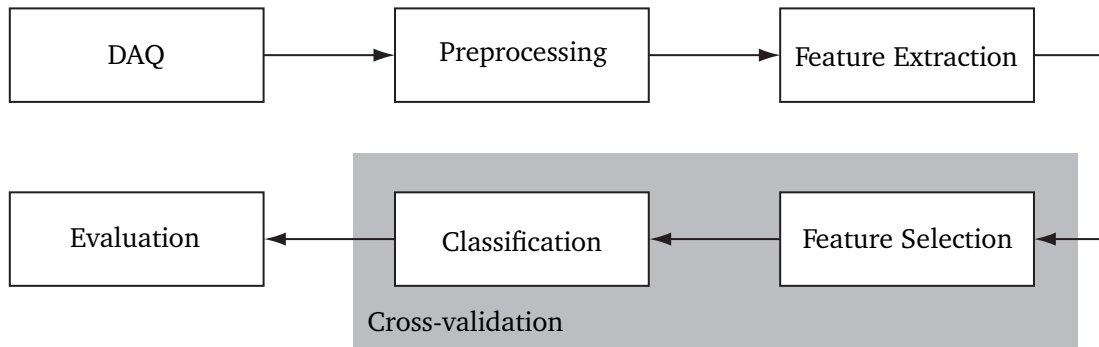
Figure 2.4: Common machine learning processing pipeline for emotion recognition.

2.2.2 Tools and practical issues

In an effort to standardize experimental procedures and thus, reduce uncontrolled factors, we developed tools that automate necessary steps and randomize conditions. First, to facilitate selection of IAPS pictures for stimuli from the database (and avoid any bias from hand-selection), an interface including a GUI was developed to automatically select most appropriate pictures corresponding to an affective state. Second, the experimental protocol was embedded in an induction software tool, which automated the sequential steps involved (e.g. showing pictures, taking user input from the SAM test, execute predefined waiting periods), randomized the order in which conditions and pictures were presented, and provided synchronization signals to the recording devices. Some details on both tools can be found in Appendix A.1.

2.3 Emotion recognition

In this section, we introduce the machine learning process that is followed by most work in this area, especially when static methods for emotion recognition are applied. Since the remaining chapters of this thesis focus on more specialized topics, we introduce important general work in the field of emotion recognition from physiological signals here. Compared to other modalities like facial expressions, physiological signals have the advantage that they cannot be easily manipulated by the subject. They can be divided into brain signals and peripheral signals like heart activity, pulse, temperature, and galvanic skin response (GSR).

The common processing pipeline in emotion recognition is visualized in Fig. 2.4 and usually includes data acquisition (DAQ), preprocessing, feature extraction, feature selection, classification or regression, and evaluation. They are considered each in turn in the following; additionally, we add other relevant issues in emotion recognition at the end of this section.

2.3.1 Data acquisition (DAQ)

For the recording of physiological signals, commercial solutions are readily available that allow for a seamless integration of sensors and amplifiers into the setup. For the experimental work described in this theses, we deployed a g.tec system which is connected to the recording

PC via a USB interface [47]. In general, brain signals are measurable through various invasive and non-invasive methods. The latter include, for example, electroencephalograms (EEG), near-infrared spectroscopy (NIRS), or functional magnetic resonance imaging (fMRI) [48]. They differ in their temporal and spacial resolution (see also [7]). Invasive methods for brain signals are typically not considered for emotion recognition, since it implies a serious surgical intervention that is out of proportion to the benefits that can be gained for a subject at the current level of research. Peripheral signals are accessible through signal-specific sensors, for example, heart activity is recorded via electrocardiography (ECG), GSR is measured by two electrodes attached to the finger tips, and optical sensors are used for blood flow and pulse measurements.

During the DAQ phase, it is important to assure appropriate synchronization of the presented stimuli and recorded data to make segmentation, as covered in the next section, feasible.

2.3.2 Preprocessing

Preprocessing steps refer to actions like segmentation and cleaning of data. In segmentation, the parts of the data that is of no interest, e.g. recording phases during times of relaxation in the experiment, are discarded and data of interest is separately labeled and stored for further processing.

Cleaning can involve band or notch filters to reduce noise or the application of more sensor-specific and often elaborate methods for artifact removal, as it is common especially for EEG signals. An exemplary tool to identify and remove different kinds of artifacts, e.g. eye-blinks, eye-movements, drift, and generic discontinuities, is the Independent Component Analysis (ICA) based plug-in ADJUST [49] for the MATLAB Toolbox EEGLAB [50].

Often, data is down-sampled in order to relax computational cost of further processing.

2.3.3 Feature extraction and selection

A multitude of features from physiological signals have been suggested. They are computed from time and frequency domain and are often motivated by neuro-scientific findings or somatic functions. For overviews of features from both peripheral physiological signals and EEG signals, the reader is referred to [51] and [201], respectively.

When a large amount of features is computed, feature selection techniques are applied to reduce the feature space and avoid overfitting. Since this is a pressing issue especially in the field of emotion recognition from EEG, we devote Chapter 4 to this topic and therefore keep this section brief.

2.3.4 Classification

Depending on the emotion model applied, either classification or regression methods are applied to predict the emotional state of unseen physiological data. When predicting discrete emotion classes, commonly used classifiers that have been developed in machine learning include Discriminant Analysis [52], Support Vector Machine (SVM) [53], k-Nearest Neighbor (kNN) [52], or Neural Networks [54]. In dimensional space, regression techniques are

applied including, for example, linear regression in its regularized form, i.e. ridge regression [55]. Feature Selection and Classification is usually carried out using cross-validation, i.e. a part of the data is set aside for testing after feature selection and training of the classifier is carried out using the rest of the data [52].

2.3.5 Evaluation

In classification tasks, evaluation is usually done by reporting the accuracy, i.e. the percentage of predicting the correct emotion class. Often results are compared against baselines using statistical testing to detect significant differences. In regression tasks, i.e. when predictions are made in one or more continuous dimensions, alternative measures are used. Since we review measures for emotion recognition on continuous scales in Section 6.2, we refrain from giving a full overview here.

Often, results have a complex structure and are not simply measurable by a scalar number. Thus, particularly for brain activity studies, sophisticated visualizations are necessary to convincingly convey results to the reader.

2.3.6 Related issues

Some related topics not directly addressed in this thesis but relevant in the field of emotion recognition include:

- Multi-modal fusion: the fusion of emotion markers from multiple modalities can be achieved at either data, feature, or classifier level and is believed to compensate for weaknesses of one modality by adding another modality [56]–[58].

- Online recognition: since data collection from subjects is very time-consuming, most current studies focus on offline analysis of the data. The ultimate goal of emotion recognition being online applications, some attempts for online processing of data have been made [59]–[64]. In Part II, we deal with the related topic of dynamic emotion recognition, which does not address online processing issues, but the continuous estimation of emotion over time.

- Real-world setting: there are still major challenges for transferring findings from laboratory conditions to real-world applications. For example, motion and daily activities affect the physiological measurements and can deteriorate their quality for emotion recognition [65]. Simultaneous advancement of sensor technology and knowledge on emotion recognition gained from controlled experimental conditions is necessary to overcome these problems.

Part I

Static emotion recognition

3 A comparison of feature selection methods detecting interacting features

Summary. *The implications of interacting features in feature selection problems, such as they are likely to occur in features from physiological signals, have not yet been definitely clarified. In this chapter, we aim to improve the understanding of this topic by studying issues related to interacting features. Herein, fully controllable experimental conditions are necessary to be able to draw meaningful conclusions. After giving an overview on feature selection filter methods and presenting selected state-of-the-art feature selection methods in detail, the main steps presented in this chapter include*

- *the discussion of four types of feature interactions in multi-class problems,*

- *the design of a flexible synthetic database comprising these interactions,*

- *a study comparing selected feature selection methods w.r.t. their capabilities of detecting and exploiting feature interactions.*

In emotion recognition from physiological signals, especially from EEG, there are a vast amount of possible features and electrode locations. But it is not clear, which features and electrodes are of particular interest and most relevant for emotion recognition. Feature Selection (FS) methods can help detect suitable features as we will show in the study given in Chapter 4. In the complex and interconnected system of the human body it is likely that features interact in ways that certain combinations of features perform better than each feature by itself. However, research on interacting features in the context of feature selection is limited and a better understanding of the same would be beneficial in several aspects, particularly w.r.t. the study following this chapter.

3.1 Problem statement and approach

Most work on FS for interacting features use real-world datasets, e.g. from the *UCI Machine Learning Repository* [66], to test and evaluate algorithms [67]–[69]. This approach has the disadvantage that experimental conditions, for example the type and degree of interaction, cannot be systematically controlled and the expressiveness of conclusions might, thus, be limited. Only few studies make an effort to additionally design artificial datasets with specific build-in interactions. Zhu et al. report such a multi-class dataset containing one type of interaction to evaluate variations of their proposed FS method [69]. Other synthetic datasets available today for comparing FS methods, e.g. the famous MONK problem [70], are often limited to binary classes and features [71], [72]. Yet, this is not suitable for the problem of

emotion recognition which commonly deals with continuous features and multiple classes.

The task is to develop a systematic analysis to improve the understanding of how various types of interacting features can influence results, e.g. classification performed on the basis of certain features, and which FS methods are able to detect and exploit them. Experimental analysis should therefore be carried out on an artificial dataset where parameters can be directly controlled to gain a better understanding of the challenges of detecting interactions as well as the strength and weaknesses of different FS methods. An improved understanding of the underlying mechanisms can lead to a better choice of FS methods given a problem, which in turn can lead to improved results, and can even allow better interpretability and conclusions of results from real-world data. In this chapter, we define and discuss various types of feature interactions and carry out a study on a synthetic dataset. In this, we chose an academic example under some common assumptions about the data and compare a selection of state-of-the-art FS methods on this dataset with the question in mind, under which circumstances interactions can be detected and exploited. As the number of available samples is often an issue in real-world domains, we share the concern of Belanche and González [71] that different sample sizes may lead to different results and, hence, include this parameter in our study as well.

The remainder of the chapter is organized as follows: Section 3.2 gives an overview on feature selection methods. We define various types of feature interactions in Sec. 3.3. An academic example to study interaction is reported in Sec. 3.4. This starts with the design of a synthetic dataset on which simulations are run in Sec. 3.4.1. Section 3.4.2 gives details on the simulations and Section 3.4.3 analyzes the results. Conclusions are presented in Sec. 3.5.

3.2 Overview of feature selection methods

With the abundant production of data, dimensionality reduction has become increasingly important for classification tasks. The *curse of dimensionality* – as termed by R. Bellman [73] – points to the issue that arises in high-dimensional space when two points (from the same class) that are close to each other in low-dimensional space are likely to have a large distance to each other in high-dimensional space. This counter-intuitive phenomenon requires us to reduce dimensionality in order to improve the performance of the classifier.

Generally, two options are available. On the one hand, feature transform techniques like Principal Component Analysis (PCA) [74], Multidimensional Scaling (MDS) [75], or Linear Discriminant Analysis (LDA) [52], which utilize orthogonal transformations, are applied. That is, a projection of the data onto lower-dimensional space which still requires the measurement of all features. On the other hand, feature selection techniques can be employed which, in contrast, search the feature space for a subset of features that are most important. The overarching goals of FS is to improve classification performance, reduce measurement costs, and to speed up the computational process by selecting most relevant and important features for classification as well as removing irrelevant and redundant features.

We define the matrix of all features as $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_F]$, where $\mathbf{x}_i$ is the vector of all samples of one feature and F is the number of features. An instance of a feature is denoted x. The target vector containing the class information, denoted $c = 1, \ldots, C$ where C is the

number of classes of the given problem, is written as **y** and has the same dimensions as a feature $\mathbf{x}_i$, i.e. each sample belongs to one class. Given the set of all features $\mathscr{X}$, the problem of feature selection can be defined as finding the subset of z features $\mathscr{S}$ which shows the 'best' predictive characteristics w.r.t. estimating the targets **y**.

Several reviews on feature selection methods are available due to its gained importance in various fields like gene expression array analysis, text categorization, and other data mining applications. First attempts to structure different FS methods are given by Dash and Liu [76] and Blum and Langley [77], both in 1997. When the number of features used in studies increased severely from tens into the hundreds and thousands, Guyon and Elisseff surveyed FS methods in order of their complexity [78]. A recent and extensive review is produced by Saeys et al. who focused on the application in the field of bioinformatics [79]. Further, Liu and Motoda provide recent developments to extend Feature Selection (FS) algorithms, for example, making it efficient on large scale datasets, and give examples of application in text classification and bioinformatics [80]. More specialized reviews focusing on subfields like FS search methods [81] or FS for synthetic problems [71] have appeared after this.

Feature Selection methods can generally be divided into filter, wrapper, and embedded methods [77]. While wrapper methods select features based on interaction with a classifier, i.e. an underlying model, embedded methods combine the search of feature subsets and model hypothesis. Filter methods are model-independent since they evaluate the intrinsic properties of the data. This as well as their often simple implementation makes filter methods very popular. Another advantage of filters is that they usually require less computational power than wrappers and are, hence, more suitable for large data sets containing many features. The focus in this thesis lies on filter methods, which have been reviewed w.r.t. their application in gene expression microarray analysis by Lazar et al. [82]. Filter techniques can be grouped into univariate and multivariate methods. They rely on a different strategy to select features. In the following, we introduce each strategy and present a selection of state-of-the-art FS methods frequently used today as well as statistical measures for multivariate FS investigated within the scope of this thesis.

3.2.1 Univariate filter methods

Univariate filter methods evaluate each feature individually according to a specific criterion and return a ranking of features. Such criteria are either parametric, i.e. they make some assumptions about the data's distribution, or non-parametric. In the FS process, each feature is evaluated individually and assigned the value according to the selected criterion. The values are then sorted by rank and the top z features are returned as the selected subset $\mathscr{S}$. Often used criteria are correlation criteria and related test statistics [78], mutual information between a feature $\mathbf{x}_i$ and the class label **y** [83], and ranking criteria like the Wilcoxon rank sum [84]. An example for a non-parametric criteria is proposed in the *ReliefF* algorithm given in more detail below, followed by a statistical criteria from the parametric group.

ReliefF

This univariate technique is an extension of the *Relief* algorithm, which uses a subsample of all instances to adjust weights of each feature depending on their ability to discriminate between two classes [85]. Feature selection is done by estimating a quality weight W_i for each feature $\mathbf{x}_i$, $i = 1, \dots, F$ based on the difference diff($\cdot$) to the nearest hit $\mathbf{x}_H$ and nearest miss $\mathbf{x}_M$ of m randomly selected instances $\mathbf{x}_j$:

$$W_i = W_i - \text{diff}(i, \mathbf{x}_j, \mathbf{x}_H)/m + \text{diff}(i, \mathbf{x}_j, \mathbf{x}_M)/m \ . \tag{3.1}$$

To overcome the high sensitivity to noise, *ReliefF* uses k nearest hits/misses. Multiple classes are accounted for by searching for the k nearest instances from each class weighted by their prior probabilities. A MATLAB implementation is readily provided in the statistics toolbox.

Statistical criteria

Several test statistics have been proposed for feature selection [79]. Here, we introduce the effect size (ES) measure f^2 from Analysis of Variance (ANOVA) that draws on the difference between within-class and between-class variance of a feature. Its relation to Bayes classification – and therefore its suitability as an evaluation function – has been investigated and discussed in [207], [86]. The ES f^2 criterion is closely related to the frequently used Fisher Discriminant Ratio (FDR, e.g. [64], [87]), the F statistic and other similar criteria as also discussed by Guyon and Elisseeff [78]. An advantage of using the effect size is that it does not depend on the sample size. We choose it over the other criteria for consistency in effect size-based feature selection presented for the multivariate FS methods in the next section.

Cohen's effect size f^2 is a generalization to more than two classes of Cohen's d:

$$d = \left| \frac{\mu_\mathbf{x}^1 - \mu_\mathbf{x}^2}{\sigma} \right| \ , \tag{3.2}$$

which is commonly reported for the statistical t-test [88]. The spread of the class means $\mu_\mathbf{x}^i$, computed from the samples belonging to class i, in the numerator is represented as a standard deviation σ_m. The denominator remains the pooled standard deviation σ of the populations involved. Thus,

$$f = \frac{\sigma_m}{\sigma} \ , \text{ where } \ \sigma_m = \sqrt{\frac{\sum_{i=1}^{C} (\mu_\mathbf{x}^i - \mu_\mathbf{x})^2}{C}} \tag{3.3}$$

for equal sample size N per class. Here, $\mu_\mathbf{x}$ is the overall mean of a feature $\mathbf{x}$. The number of classes is denoted by C.

3.2.2 Multivariate filter methods

While univariate filters output a ranking of the feature set, multivariate filters assess the suitability of a whole subset at once and return the 'best' subset of z features. Following Lazar et al., this requires three main steps to be carried out [82]:

1. Define a cost function to optimize.

2. Choose an optimization algorithm, i.e. search method, to generate subsets.

3. Validate the selected subset of features.

The most popular cost functions, i.e. the counterparts to the univariate criteria, can be grouped in Correlation Based Feature Selection (CBFS) criteria [89], Mutual Information based criteria [90] including the well-known mRMR method [91] explained below, and Conditional Entropy based approaches as suggested by Koller and Sahami [92] with the most recent extension presented by Mamitsuka [93]. Debout has revised these methods in more detail [196]. Further, statistical criteria from multivariate ANOVA can be employed, which we also introduce below.

Min-Redundancy-Max-Relevance (mRMR)

The most famous method using mutual information to characterize the suitability of a feature subset is called minimal-Redundancy-Maximal-Relevance (mRMR) and was developed by Ding and Peng [91], [94]. Mutual Information between two random variables $\mathbf{x}$ and $\mathbf{y}$ is defined as

$$I(\mathbf{x};\mathbf{y}) = \iint p(\mathbf{x},\mathbf{y}) \log \frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{x})p(\mathbf{y})} d\mathbf{x}d\mathbf{y} \, , \tag{3.4}$$

where $p(\mathbf{x})$ and $p(\mathbf{y})$ are the marginal probability density functions of $\mathbf{x}$ and $\mathbf{y}$, respectively, and $p(\mathbf{x},\mathbf{y})$ is their joint probability distribution. If $I(\mathbf{x};\mathbf{y})$ equals zero, the two random variables $\mathbf{x}$ and $\mathbf{y}$ are statistically independent. The multivariate mRMR method aims to optimize two criteria simultaneously:

1. Maximal-relevance criterion D, which means to maximize average mutual information $I(\mathbf{x}_i;\mathbf{y})$ between each feature $\mathbf{x}_i$ and the target vector $\mathbf{y}$.

2. Minimum-redundancy criterion R, which means to minimize the average mutual information $I(\mathbf{x}_i;\mathbf{x}_j)$ between two features.

The algorithm finds near-optimal features using forward selection. Given an already chosen set $\mathscr{S}_z$ of z features, the next feature is selected by maximizing the combined criterion $D-R$:

$$\max_{\mathbf{x}_j \in \mathscr{X}-\mathscr{S}_z} \left[I(\mathbf{x}_j;\mathbf{y}) - \frac{1}{z} \sum_{\mathbf{x}_i \in \mathscr{S}_z} I(\mathbf{x}_j;\mathbf{x}_i) \right] \, . \tag{3.5}$$

A necessary step to use the toolbox provided by Peng et al. [95] is to discretize the data beforehand.

Statistical criteria

Similar to the aforementioned f^2, multivariate effect sizes exist that can be used for FS. Multivariate effect size measures from corresponding multivariate ANOVA (MANOVA) in statistics are based on the eigenvalues λ_i of the generalized eigenproblem

$$\mathbf{H}\boldsymbol{\nu} = \lambda \boldsymbol{\nu}\mathbf{E} \, , \tag{3.6}$$

Tabular 3.1: MANOVA statistics and effect size measures

Name	Statistic	Effect size
Wilk's Λ	$\Lambda = \prod_{i=1}^{q}\left(\dfrac{1}{1+\lambda_i}\right)$	$\tau^2 = 1 - \Lambda^{1/q}$
Pillai's trace	$U = \sum_{i=1}^{q}\left(\dfrac{\lambda_i}{1+\lambda_i}\right)$	$\eta_{\mathrm{PB}}^2 = \dfrac{U}{q}$
Roy's θ	$\theta = \dfrac{\lambda_i}{1+\lambda_i}$	
Hotelling criterion	$V = \sum_{i=1}^{q}\lambda_i$	$\zeta^2 = \dfrac{V}{q+V}$

where **H** and **E** are the between and within sum of square and cross product (SSCP) matrices, respectively [96]. Since solving (3.6) requires the computation of the inverse of **E**, features have to be uncorrelated. Four common statistics and corresponding effect sizes listed in Table 3.1 are computed from the eigenvalues λ_i.

The effect size based on Wilk's Λ, which has been applied for feature selection in [97], is defined as

$$\tau^2 = 1 - \Lambda^{1/q}, \text{ where } \Lambda = \prod_{i=1}^{q}\left(\frac{1}{1+\lambda_i}\right) \tag{3.7}$$

and simplifies to f^2 in the univariate case. The variable $q = \min(F, C-1)$, where F is the number of features and C the number of classes. It adjusts for the degrees of freedom and the number of eigenvalues used to compute the statistic. We abbreviate this FS method based on Wilk's Λ by writing ES Λ in the following.

We further propose the use of Roy's statistic θ for FS, which considers the largest eigenvalue only, i.e. the explained variance of the first underlying construct and is defined as

$$\theta = \frac{\lambda_1}{1+\lambda_1} \ . \tag{3.8}$$

Since it does not need to be normalized, θ can be used directly as a measure of effect, and we denote FS based on this criterion as ES θ. For large eigenvalues, both statistics converge to one. An in-depth analysis of MANOVA effect sizes and their mathematical derivation is offered by Huberty and Olejnik [96, pp. 61ff.].

Search methods. As indicated above, the second main step after deciding for a cost function is to select an optimization algorithm in form of a search method, which generates the next feature subset $\mathscr{S}$ to be evaluated. Besides Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS), which are the most common search methods in FS, several others including randomized subset generation exist. Many of these search methods have been reviewed by Somol et al. [81] and can provide useful extensions to FS by combining them with known cost functions as the ones presented above. This can bring about advantages in terms of speed, especially for large datasets, as well as avoiding local minima of the cost function, which is a well known problem of SFS [80]. A study on different search

methods, which might be worth looking into in more detail as an extension of this work, is reported in [195].

Here, with the aim of comparing the FS methods introduced above, we chose SFS as search method for the statistical criteria associated with ES Λ and ES θ, as it is also used in the mRMR algorithm. That means, the feature that returns the highest ES measure when added to the already chosen set of features $\mathscr{S}$ is selected.

3.3 Interactions of features

Generally, *interacting features* are features that carry conditionally relevant or redundant information w.r.t. the target class. In this section, we present four examples that will be studied in simulation later. Most helpful lines of literature on the broader aspects of interactions are probably given by Jakulin and Bratko [67], [98] as well as Zhao and Liu [68]. Jakulin and Bratko introduced the definitions of *positive* and *negative* interactions, where the former covers unexpected synergies that increase classification performance while the latter points to the presence of redundant information [67]. Positive interaction can be subdivided into several subgroups further specifying the nature or source of interaction. Within this chapter, we focus on linear interactions. Non-linear interaction are commonly dealt with by including non-linear terms of features in the extraction step.

Guyon and Elisseeff discussed three exemplary forms of possible interactions that expose limitations of univariate ranking methods for dichotomous problems [78]. Here, we generalize these to multi-class problems with equally distributed class means. We further add a case of unequally distributed class means, which can be interpreted as a generalization of the XOR problem for more than two classes. Figure 3.1 shows four scenarios of interaction assuming data from Gaussian distributions, which are discussed in the following.

I.i.d. but not truly redundant. The trivial case of positive interaction is given in Fig. 3.1a. While features are independently and identically distributed (i.i.d.), they are not truly redundant. As Guyon and Elisseeff pointed out, using both features improves separation by a factor $\sqrt{2}$ [78].

Intra-class covariance. Figure 3.1b and 3.1c display positive and negative interactions, respectively, which result from non-zero covariance between the two features. The latter occurs when the covariance is large in the same direction as the class means are distributed (Fig. 3.1c). In this case, feature 2 is said to be redundant given feature 1. When the covariance is large in a direction different from the class means, even a feature that has no separability by itself (here, feature 1 in Fig. 3.1b) can become very important in combination with another feature (here, feature 2).

Unevenly distributed means. The three cases mentioned so far all assume evenly distributed means. When this assumption is not valid - as it is likely in multi-class problems - interaction can occur, for example, in the form plotted in Fig. 3.1d. Here, feature 1 separates class 1 and 2 from class 3, but cannot distinguish between class 1 and 2. Combined with

feature 2, however, all three classes can be separated. Zhu et al. described this type of interaction as Partially Class Relevant (PCR) features [69].

Certainly, combinations of the presented interaction types can be imagined and might further complicate the detection of optimal features. For an academic example of comparing the performance of FS methods, we will use the interactions introduced above to design an artificial dataset on which we carry out simulations described in the next section.
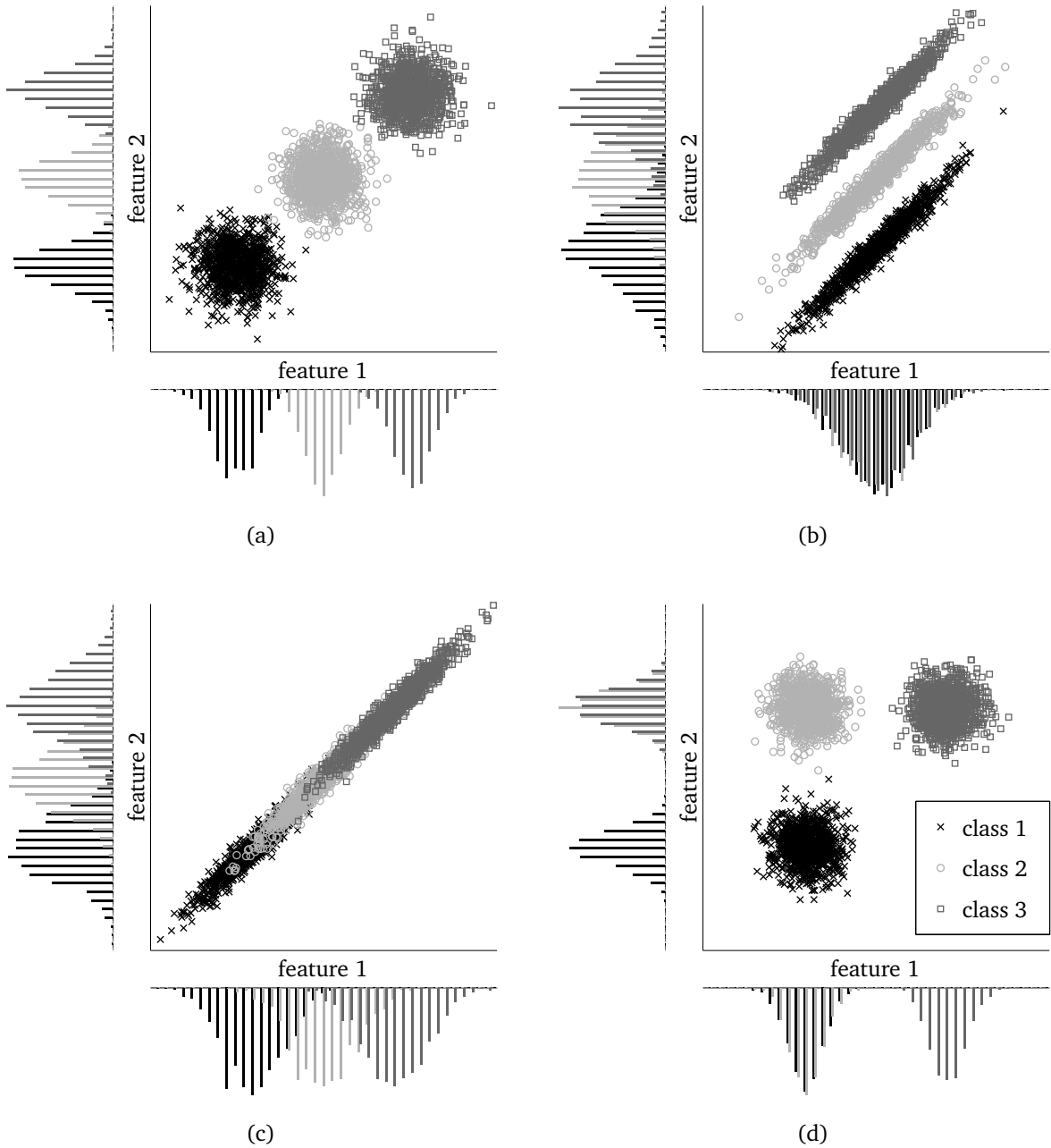

Figure 3.1: Types of interaction: (a) i.i.d. but not truly redundant, (b) positive intra-class covariance, (c) negative intra-class covariance, (d) unequally distributed means.

3.4 An academic example

In order to study how the presence of different interaction types influence feature selection using both univariate and multivariate FS methods, a structured dataset has been designed and generated. On this we simulated feature selection systematically to find out which types of interactions can be detected by which FS method, where multivariate methods are needed, and how the sample size of the data influences results. An academic example is chosen here over a real-world dataset to be able to properly test and discuss desired variations of properties of the data and, thus, arrive at more meaningful conclusions. Findings from this chapter can contribute to a better interpretation of the analysis of a real-world dataset presented in Chapter 4. This section describes the dataset generated and presents the process as well as the outcomes of the simulations.

3.4.1 Database

Reviews on existing synthetic datasets can be found in [71], [99]. Of those, however, none represents multiple types of interactions as introduced in Sec. 3.3, which is why we present our method for data generation here. In designing the four datasets according to the four interaction types under investigation, we make the following assumptions, similar to those made by Zhu et al. [69]:

- equal sample size N per class,

- data drawn from a Gaussian distribution with equal covariance matrix Σ^c per class,

- except for the last interaction type, we assume means to be equally distributed for each feature.

Each set of features $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_F]$ consists of F features $\mathbf{x}_i, i = 1, \ldots, F$ with N samples from C classes, where $\mathbf{X}^c = [\mathbf{x}_1^c, \ldots, \mathbf{x}_F^c]$ denotes the samples of all features belonging to class $c = 1, \ldots, C$. Features for each class are generated from a multivariate normal distribution with mean $\boldsymbol{\mu}^c$ and covariance matrix Σ^c using the MATLAB function `mvnrnd`, such that

$$\mathbf{X}^c \sim \mathcal{N}(\boldsymbol{\mu}^c, \Sigma^c) . \tag{3.9}$$

Each dataset consists of a total of six features, two of which, namely $\mathbf{x}_1$ and $\mathbf{x}_2$, are interacting such that they achieve the best accuracy of all features when combined, two 'distracting' features, $\mathbf{x}_3$ and $\mathbf{x}_4$, which are i.i.d. and that each when considered alone separates all classes to a moderate degree, and finally two irrelevant features, $\mathbf{x}_5$ and $\mathbf{x}_6$, from which no class information can be gained. The parameters $\boldsymbol{\mu}^c$ and Σ^c are chosen to fulfill the characteristics of each of the four types of interactions and are given below. The range of means has been selected such that the task of feature selection is sufficiently difficult to expose differences in the FS methods. Of the selected range of means, the dataset generation is repeated 10 times in order to balance the results for unlikely outliers. Further, we generated sets for sample sizes $N = \{10, 20, 50, 100, 1000\}$ per class for each type of interaction. Since we usually deal with multiple classes in emotion recognition (especially when taking on discrete emotion models), we designed the dataset for four classes.

I.i.d. but not truly redundant. In the first case, features are i.i.d., i.e. Σ^c is a diagonal matrix with all non-zero entries equal to one. The basis mean μ^c is defined as a function of the variable step size s such that

$$\mu^c = (c-1)s \, , \tag{3.10}$$

where $s = \{.6, .7\}$ for the current interaction type. The means vector $\boldsymbol{\mu}^c$ of all features of one class is then chosen as

$$\boldsymbol{\mu}^c = [\mu^c, \mu^c, a \cdot \mu^c, a \cdot \mu^c, 0, 0] \, , \tag{3.11}$$

such that the resulting features of each entry of $\boldsymbol{\mu}^c$ correspond to the six features explained above. The scaling factor a is set to .9 in this case.

Intra-class covariance. In the case of positive and negative interaction through intra-class covariance, non-zero off-diagonal entries are added to the matrix Σ^c for the interacting features $\mathbf{x}_1$ and $\mathbf{x}_2$, i.e. it takes the form

$$\Sigma^c = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & 0 & 0 & 0 & 0 \\ \sigma_{21}^2 & \sigma_{22}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} . \tag{3.12}$$

For the case of positive intra-class interaction, we set $\sigma_{12}^2 = \sigma_{21}^2 = -.8$ and $\sigma_{11}^2 = \sigma_{22}^2 = 1$. For the case of negative intra-class interaction, we set $\sigma_{12}^2 = \sigma_{21}^2 = .7$ and $\sigma_{11}^2 = \sigma_{22}^2 = 1.1$. The means $\boldsymbol{\mu}^c$ are the same as given in (3.10) and (3.11) for the first type of interaction, with the step size chosen as $s = \{1.2, 1.3, 1.4\}$ and a scaling factor of $a = .8$ and $a = .9$ for positive and negative interaction, respectively.

Unevenly distributed means. Last, we designed a case for feature interaction resulting from unevenly distributed means. In this academic example, we set the covariance matrix to be diagonal with all non-zero entries equal to one. The means of the first class is defined as

$$\boldsymbol{\mu}^1 = \left[0, 0, a \cdot \mu^1, a \cdot \mu^1, 0, 0\right] \, , \tag{3.13}$$

and that of the remaining three classes as

$$\boldsymbol{\mu}^{c'} = \left[\mu^{c'-1}, s, a \cdot \mu^{c'}, a \cdot \mu^{c'}, 0, 0\right], \text{ for } c' = 2, 3, 4 \, . \tag{3.14}$$

This way, probability distributions of class 1 and 2 are overlapping for feature $\mathbf{x}_1$, but all other classes can be discriminated between using this feature. Feature $\mathbf{x}_2$ compliments $\mathbf{x}_1$ in that it can only separate class 1 from all others. The step size $s = \{2.5, 2.6, 2.7\}$ in this case and the scaling factor $a = .45$.

3.4.2 Simulations

The process to answer the questions stated above follows the common machine learning approach applied after feature extraction, i.e. feature selection, classification, and evaluation [55], [100]. The dataset is generated once and stored so that in simulation identical data were presented to each FS method.

Simulations were performed for data of each interaction type and different numbers of samples N per class. A 10-fold cross validation was applied in which 9 folds were used for FS and subsequent training of the classifier, while the remaining fold was used for testing. Each fold was used once for testing and the results were averaged over all 10 folds. For the required discretization of the data for FS method mRMR, we chose 20 levels. As a classifier, linear discriminant analysis (LDA) with pooled estimate of covariance was implemented. We chose LDA, since by design the data was linearly separable and thus, it was not necessary to apply more sophisticated classifiers such as Support Vector Machine (SVM) or non-linear discriminant analysis to evaluate results of FS. Ultimately, the possible gain in classification accuracy when using the most suitable FS method for the task at hand is desirable. Thus, the accuracy of each set of features is adopted as an evaluation measures. Results are presented in the next section.

3.4.3 Results and discussion

Figures 3.2 and 3.3 show the results gained from simulations: The FS methods are grouped by the number of samples N per class and are spread along the x-axis. The y-axis gives the mean and variance of the achieved accuracy for each simulated case. A star above a bar marks that the best method of the group (shaded dark gray) performs significantly higher than the method above which the star is plotted. Statistical tests are carried out as pair-wise t-tests with Bonferroni correction at a global α-level of .05.

I.i.d. but not truly redundant. The first type of dataset is used as a baseline, since both univariate and multivariate methods should be able to detect the best features as long as they do not assume one of the features to be redundant and throw it out. As can be seen in Fig. 3.2a, this holds for large numbers of samples as expected with a low variance of accuracy which increases naturally for smaller sample sizes.

For small numbers of samples, however, univariate FS methods show a generally lower variance of accuracies, i.e. appear to be more stable in their selection of features. This can be explained by a probable overfitting of complex structure to the data by multivariate methods due to outliers. Noteworthy, this effect can also be observed between the multivariate methods ES Λ and ES θ: although the latter only considers the largest eigenvalue, i.e. the first underlying construct, and thus generally ignoring information, this method can be more robust in such cases.

Further, the mRMR method performs significantly worse than the best method, when only very few samples ($N = 10$) are available, but quickly improves performance with increased sample size.
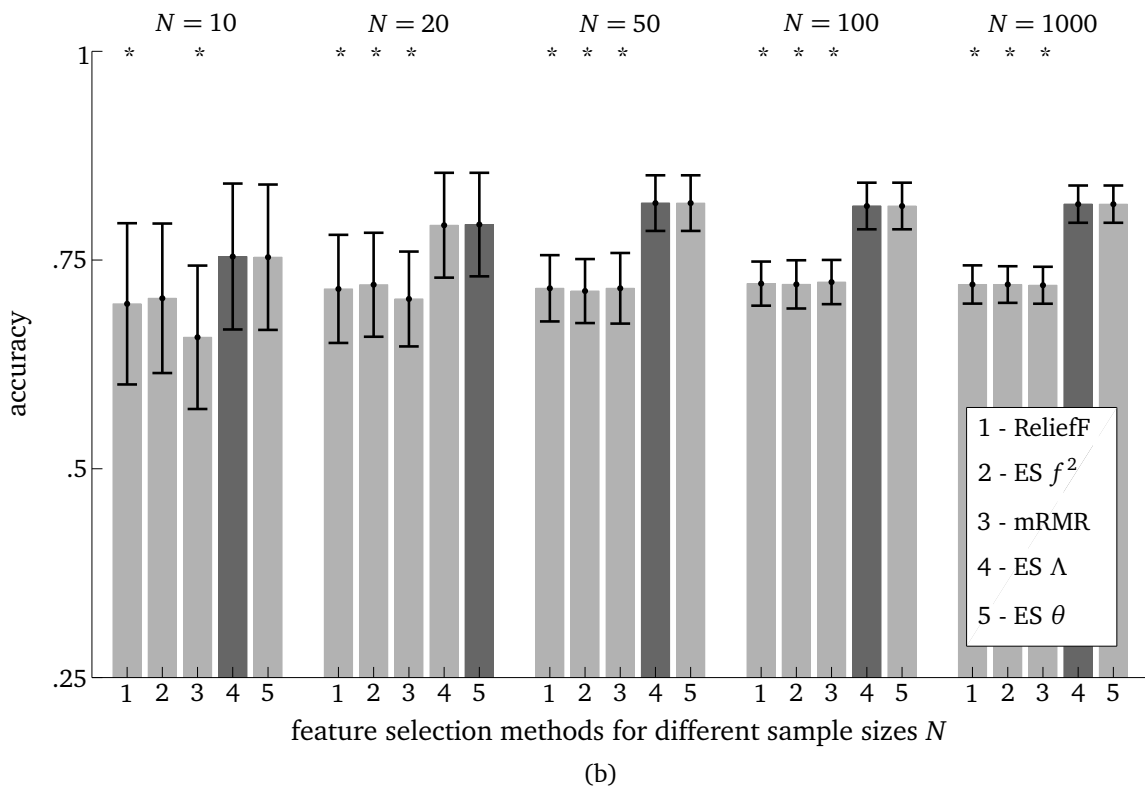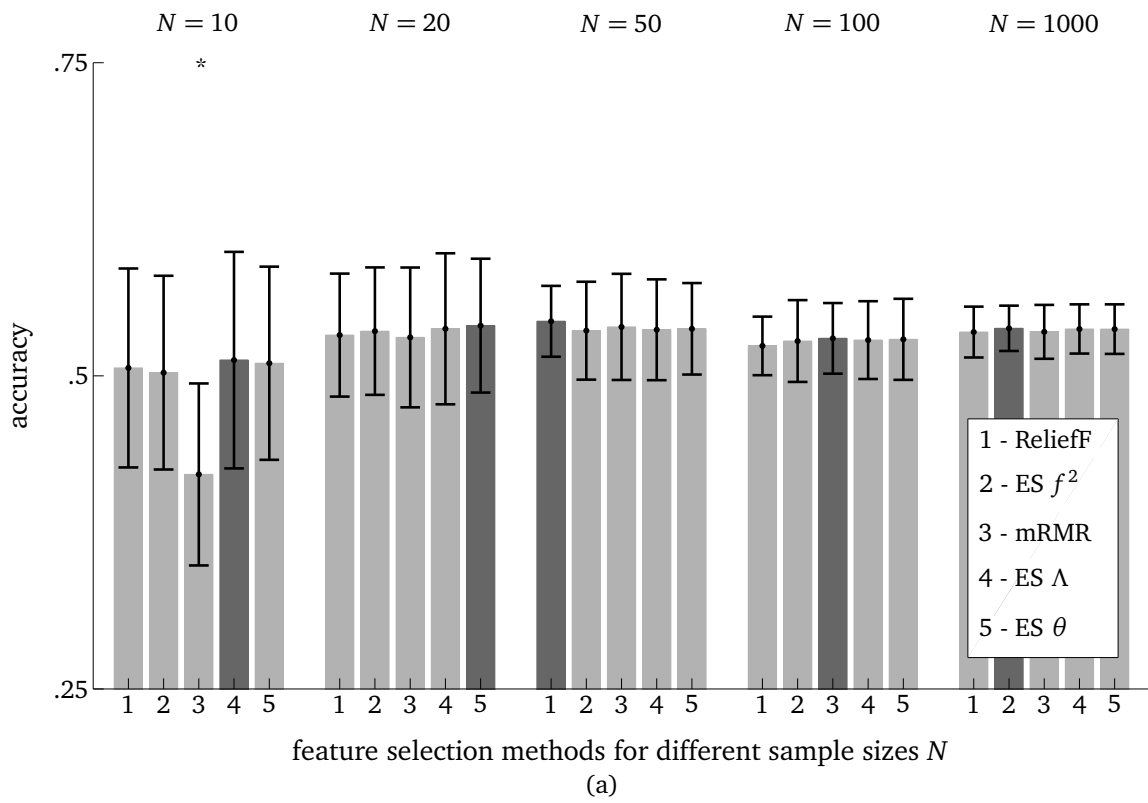
Figure 3.2: Results for interaction types (a) i.i.d. but not truly redundant, (b) positive intra-class interaction. * denotes a significant difference of the denoted bar and the best FS method (dark gray bars).

Intra-class covariance. When we introduce a positive interaction by varying the intra-class covariance matrix, the benefit of multivariate techniques starts to show (see Fig. 3.2b). For very small sample size ($N = 10$), the difference between multivariate and univariate methods is not major and the mRMR method even performs a little worse than univariate methods. But the ability of both ES Λ and ES θ to detect and exploit this interaction becomes clearer with increased sample size.

Figure 3.3a shows the results for negative interaction through intra-class covariance. All multivariate methods are able to detect the benefit of selecting one of the independently distributed features with lower univariate discriminatory power over the negatively interacting second feature. The univariate method ES f^2 cannot detect this interaction (since it ignores correlation between features when evaluating them individually), which is apparent with a significant difference compared to the best method for large numbers of samples ($N = \{50, 100, 1000\}$). ReliefF returns accuracies in between those from ES f^2 and multivariate methods. A closer look reveals, that ReliefF chooses the optimal set of features, i.e. one of the features $\mathbf{x}_{1/2}$ in combination with one of the features $\mathbf{x}_{3/4}$, only sometimes, and instead selects the two non-interacting feature $\mathbf{x}_3$ and $\mathbf{x}_4$ most of the time. Again, for a small number of samples per class, the simpler FS methods perform equally well as the more sophisticated multivariate methods.

Unevenly distributed means. The results given data with feature interaction arising from unevenly distributed means are plotted in Fig. 3.3b. Here, regardless of the number of samples available per class, univariate methods fail to recognize the benefits stemming from combining feature $\mathbf{x}_1$ and $\mathbf{x}_2$. Among the mulitvariate FS methods, outcomes vary strongly. While ES θ cannot detect this type of interaction, since it only considers the first eigenvalue of the problem given in (3.6), i.e. the first underlying construct, ES Λ finds this interaction even for small numbers of samples N per class. The mRMR method also finds this interaction given sufficient evidence: for small values of $N = \{10, 20\}$ it reaches significantly lower accuracies than ES Λ; for $N = \{100, 1000\}$ it shows accuracy rates comparable to ES Λ.

Overall, mRMR seems suitable for large number of samples per class, but shows problems in building an appropriate model of the data for small sample sizes at least when the types of interactions studied here are present. When sufficient samples are available, ES Λ is the best method across all studied interaction types and usually outperforms ES θ, since it can detect more complicated interactions. Further, the advantages of multivariate over univariate FS methods only become a reliable asset for sufficiently large number of data samples. This is probably due to the fact that these methods tend to overfit the data otherwise. When only few samples are available (like it is common in the domain of emotion recognition from physiological signals), the use of univariate methods can stabilize results when used in conjunction with multivariate methods, which take into account possible interactions of features.
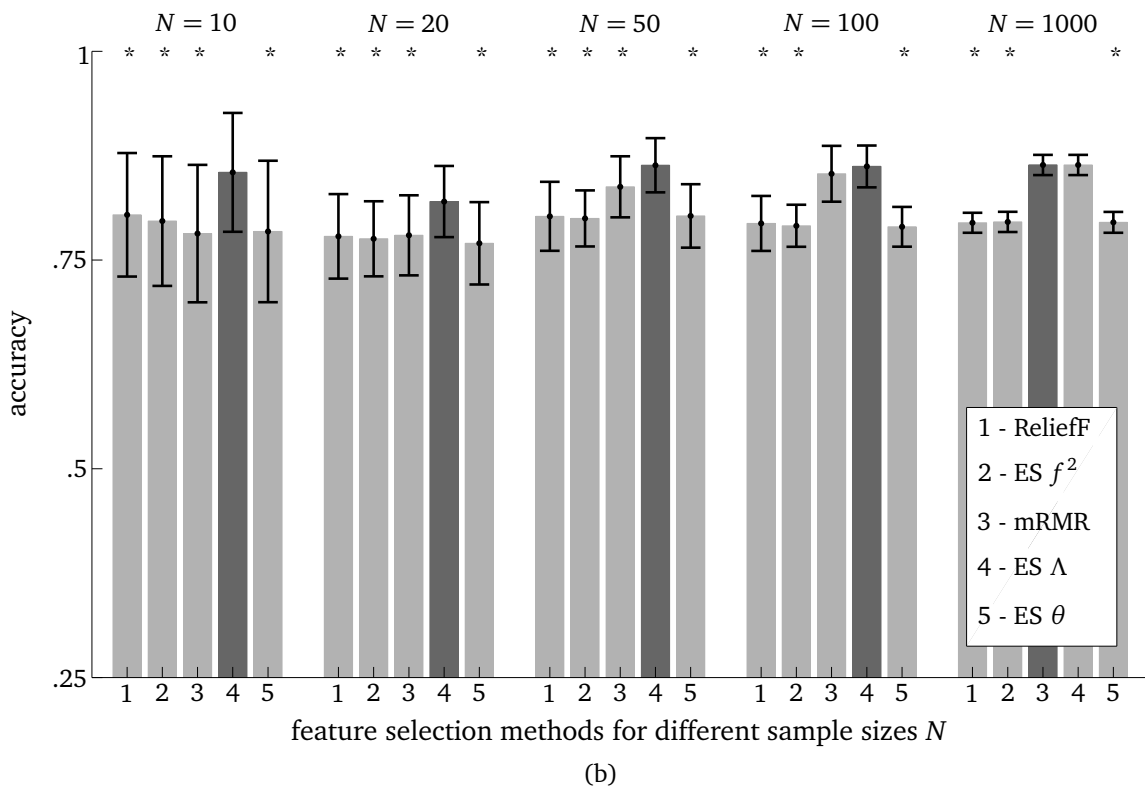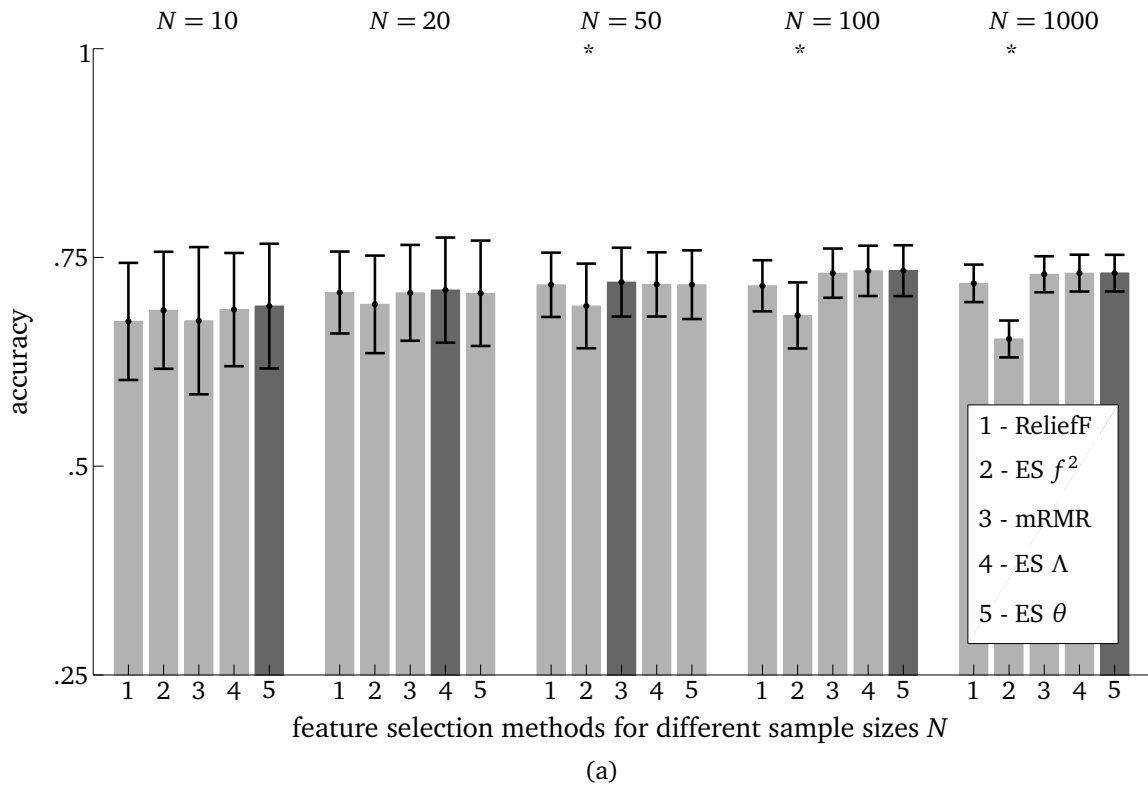
Figure 3.3: Results for interaction types (a) negative intra-class interaction (b) unequally distributed means. * denotes a significant difference of the denoted bar and the best FS method (dark gray bars).

3.5 Conclusion

In this chapter, the capabilities of both univariate and multivariate FS methods to detect interacting features in a synthetic database have been compared. We limited ourself to a simple academic example where parameters can be controlled directly to get a better understanding of the strengths and weaknesses of each method in the context of four types of interactions. If features are likely to interact, i.e. certain combinations of features perform better than each by itself, multivariate methods are needed that can detect such interaction when selecting features. The method ES Λ is best in detecting the kinds of interactions under the given assumptions here, but it is possible that this method will lack performance in other cases, justifying the commonly used alternatives. Although univariate FS methods can only judge each feature on its own, they have the advantage of being robust and less overfitting compared to multivariate FS methods especially for small numbers of samples N per class.

In accordance with the notion of feature stability as pointed out, for example, by Somol et al. [81], our results suggest that more robust conclusions can be drawn from feature selection if multiple methods are employed at the same time. To merge the results of multiple FS methods, several stability indices have been introduced in the past (e.g. [101], [102]) and we will use a similar modified index to evaluate the results of a real-world study presented in the next chapter.

While many FS methods have been benchmarked on both synthetic and real-world datasets (for an overview see [69], for example), a detailed investigation of various interaction types as attempted in this chapter was yet missing. The comparison can be extended in several ways: For example, by varying the degree of complexity of the FS task which can be achieved by increasing the number of classes, the number of interacting features, and the number of irrelevant or redundant features. Further, the performance of each method could be evaluated for combinations of interaction types. Last, as already mentioned above, FS methods can be modified by using different search strategies which could also lead to faster performance and better results. The list of FS methods could also be extended to include other schemes for multi-class FS such as the one-against-one and one-against-all schemes, as studied by Student et al. [103] in combination with Partial Least Squares (PLS) FS.

4 Feature extraction and selection from EEG

Summary. *A major limitation in emotion recognition from EEG is the unanswered question, which features and electrode locations are most suitable. In this chapter, we survey the existing literature for a comprehensive review on feature extraction techniques used in the past. On a self-recorded dataset of eleven subjects in five discrete emotional states, a systematic study applying multiple machine learning techniques for feature selection is carried out. Based on the results, we find that*

- *multivariate FS methods perform slightly better pointing towards the likely presence of interactions of features,*

- *advanced types of features outperform commonly used spectral power bands,*

- *electrode locations over parietal and centro-parietal lobes show promise.*

Interest in emotion recognition from different modalities (e.g. face, posture, motion, voice) has risen in the past decade and recently gained attention in the field of Brain-Computer-Interfaces (BCIs), which has coined the term affective BCI (aBCI) [104]. While the reliability of the ground-truth, duration and intensity of emotions remain ongoing challenges of automatic affect recognition, several attempts have been undertaken to advance the comparatively new field of emotion recognition from electroencephalography (EEG) [44]. Early work on emotion recognition from EEG goes back as far as 1997 [105], and gained more and more attention in recent years. Typically, feature extraction and electrode selection is based on neuro-scientific assumptions. In this, spectral power in various frequency bands is often associated to emotional states, as well as coherence and phase synchronization of pairs of electrodes [106]. Furthermore, frontal asymmetry in α band power as differentiator of valence levels has received a lot of attention in neurological studies [107].

Beside neuro-scientific assumptions, also advanced signal processing finds application in the field of aBCIs [108], which leads to a vast amount of possible features and makes it necessary to reduce dimensions for recognition tasks in order to avoid over-specification. Computational methods from the field of machine learning are applied to optimize the selection of features and electrodes w.r.t. achieved emotion estimation accuracy [106].

4.1 Problem statement and approach

In emotion recognition from EEG it is not generally agreed upon which features are most appropriate, and only a few works exist, which compared different features with each other. For example, Schaaff and Schultz compared two sets of features AI and AII (see Table 4.2) on a self recorded dataset of 5 subjects and three emotions [109]. Different time-frequency

techniques and frequency bands were evaluated by Hadjidimitriou and Hadjileontiadis on a dataset with 9 subjects and different levels of liking [110]. Petrantonakis and Hadjileontiadis [111] studied features in comparison to feature vectors used in [112] and [113] on a self recorded dataset of 16 subjects and six emotions.

Other studies looked into finding the best electrode positions, e.g. Wang et al. applied mRMR feature selection to find electrode positions of TOP-30 features on a self recorded dataset with 5 subjects and four emotions [114]. A channel selection method based on synchronization likelihood was tested on one subject by Ansari et al. [115]. Finally, in a previous study we investigated electrode and feature selection of selected time and frequency-domain features on a self recorded dataset of 16 subjects and five emotions [202].

However, a major limitation of these approaches is that so far only a handful of features have been compared in each study. Additionally, most studies rely on a different, yet usually small dataset. In this chapter, we aim for a comprehensive review of feature extraction methods used for affective EEG signal analysis and a systematic comparison on one dataset that allows for a qualitative evaluation of favorable features and electrodes. In doing so, we first survey techniques for feature extraction from over 30 studies and then implement and evaluate them by means of different feature selection (FS) methods on a reasonably large database of eleven subjects. Thus, the contributions of this chapter are threefold: 1) a comprehensive review of EEG features for emotion recognition, 2) a first systematic comparison of features on *one* database using multiple FS methods to arrive at more robust results when 3) comparing them to existing findings from other studies.

The remainder of this chapter is organized as follows: Section 4.2 reviews feature extraction methods used for emotion classification from EEG. Our approach to feature selection is described in Section 4.3. In Section 4.4, we present details on the recorded dataset and its evaluation. Data processing and experimental setup are presented in Section 4.5. We give results in Section 4.6 with an interpretation in Section 4.7 and end with a brief conclusion in Section 4.8.

4.2 Feature extraction

In this section, we review a wide range of features relevant for emotion recognition from EEG that have been proposed in the past and are implemented in our subsequent study unless noted otherwise. A good starting point is given in a recent study by Kim et al. [106] which we extended by several important works in the field of emotion recognition from EEG. In this, we focused on features initially proposed before the end of 2013, which is when this study was carried out. A list of relevant studies on emotion recognition from EEG to date together with the features used is given in Table 4.2. We generally distinguish features in time domain, frequency domain, and time-frequency domain. Features are typically calculated from the recorded signal of a single electrode, but also a few features combining signals of more than one electrode have been found in literature, which are listed at the end of this section.

The following notation is used: $\xi(t) \in \mathbb{R}^T$ denotes the vector of the time series of a single electrode, T is the number of time-samples in ξ. The time derivative of $\xi(t)$ is written as $\dot{\xi}(t)$. The representation of $\xi(t)$ in the frequency domain is given by $\Xi(f)$. An instance of a

feature of $\xi(t)$ is indicated as x.

4.2.1 Time domain features

Although time-domain features from EEG are not predominant, numerous approaches exist to identify characteristics of time series that vary between different emotional states.

Event Related Potentials (ERP)

Frantzidis et al. used amplitude and latency of ERPs (P100, N100, N200, P200, P300) as features in their study [116]. In an online application, however, it is difficult to detect ERPs related to emotions since the onset is usually unknown (asynchronous BCI).

Statistics of signal

Several statistical measures have been used to characterize EEG time series [63], [112], [114], [117], [118]. These are:

- power: $P_\xi = \dfrac{1}{T} \sum_{-\infty}^{\infty} \left| \xi(t) \right|^2$,

- mean: $\mu_\xi = \dfrac{1}{T} \sum_{t=1}^{T} \xi(t)$,

- standard deviation: $\sigma_\xi = \sqrt{\dfrac{1}{T} \sum_{t=1}^{T} (\xi(t) - \mu_\xi)^2}$,

- 1st difference: $\delta_\xi = \dfrac{1}{T-1} \sum_{t=1}^{T-1} \left| \xi(t+1) - \xi(t) \right|$,

- normalized 1st difference: $\bar{\delta}_\xi = \dfrac{\delta_\xi}{\sigma_\xi}$,

- 2nd difference: $\gamma_\xi = \dfrac{1}{T-2} \sum_{t=1}^{T-2} \left| \xi(t+2) - \xi(t) \right|$,

- normalized 2nd difference: $\bar{\gamma}_\xi = \dfrac{\gamma_\xi}{\sigma_\xi}$.

The normalized first difference $\bar{\delta}_\xi$ is also known as Normalized Length Density, and captures self-similarities of the EEG signal [119].

Hjorth features

Hjorth [120] developed the following features of a time series, which were used in EEG studies, e.g. [115], [121]:

- activity: $A_\xi = \dfrac{\sum_{t=1}^{T} (\xi(t) - \mu)^2}{T}$,

- mobility: $M_\xi = \sqrt{\dfrac{\text{var}(\dot{\xi}(t))}{\text{var}(\xi(t))}}$,

- complexity: $C_\xi = \dfrac{M(\dot{\xi}(t))}{M(\xi(t))}$.

Since activity is just the squared standard deviation introduced above, we omit this feature in our implementation.

Non-Stationary Index (NSI)

Kroupi et al. employed the NSI as a measure of complexity by analyzing the variation of the local average over time [119]. The normalized signal ξ is divided into small parts and the average of each segment is computed. The NSI is defined as the standard deviation of all means, where higher index values indicate more inconsistent local averages [122].

Fractal Dimension (FD)

A frequently used measure of complexity is the fractal dimension, which can be computed via several methods. For example, Sevcik's method [115], Fractal Brownian Motion [123], Box-counting [124], or Higuchi algorithm [63], [117] were employed. The latter is known to produce results closer to the theoretical FD values than Box-counting [63] and is implemented here. In order to compute the fractal dimension FD_ξ by the Higuchi algorithm [125], the time series $\xi(t)$, $t = 1, \dots, T$ is rewritten as:

$$\left\{ \xi(m), \xi(m+g), \dots, \xi\left(m + \left[\frac{T-m}{g}\right] \cdot g\right) \right\} , \tag{4.1}$$

where $[\cdot]$ denotes the Gauss notation for the floor function, $m = 1, \dots, g$ is the initial time and g is the time interval. Then, g sets are calculated by:

$$L_m(g) = \frac{T-1}{\left[\frac{T-m}{g}\right]g^2} \sum_{i=1}^{\left[\frac{T-m}{g}\right]} \left| \xi(m+ig) - \xi(m+(i-1)g) \right| . \tag{4.2}$$

The average value over g sets of $L_m(g)$, denoted $\langle L(g) \rangle$, has the following relationship with the fractal dimension FD_ξ:

$$\langle L(g) \rangle \propto g^{-FD_\xi} . \tag{4.3}$$

We obtain FD_ξ as the negative slope of the log-log plot of $\langle L(g) \rangle$ against g.

Higher Order Crossings (HOC)

Motivated to find an efficient and robust feature extraction method that captures the oscillatory pattern of EEG, Petrantonakis and Hadjileontiadis introduced HOC-based features [111]. Herein, a sequence of high-pass filters is applied to the zero-mean time series $Z(t)$:

$$\mathfrak{F}_k\{Z(t)\} = \nabla^{k-1} Z(t) , \tag{4.4}$$

Tabular 4.1: Frequency band ranges and decomposition levels of EEG signals recorded at
$f_S = 512\,\text{Hz}$

Bandwidth (Hz)	Frequency Band	Decomposition Level
1-4 Hz	δ	A6
4-8 Hz	θ	D6
8-10 Hz	slow α	D5 (8-16 Hz)
8-12 Hz	α	
12-30 Hz	β	D4 (16-32 Hz)
30-64 Hz	γ	D3 (32-64 Hz)

where ∇^k is the iteratively applied difference operator[1], and the order $k = 1, \ldots, 10$ according to [111]. The HOC sequence H_k, i.e. the resulting k features, comprises the number of zero-crossings of the filtered time series $\mathfrak{I}_k\{Z(t)\}$ by counting its sign changes.

The approach has also been successfully studied in combination with hybrid adaptive filters [126]. Preprocessing for noise reduction is implemented in this work using third level decomposition of Discrete Wavelet Transform (DWT) (see Section 4.2.3).

4.2.2 Frequency domain features

Band power

The most popular features in the context of emotion recognition from EEG are power features from different frequency bands. This assumes stationarity of the signal for the duration of a trial. The frequency band ranges of EEG signals are varying slightly between studies. Commonly they are defined as given in the first two columns of Table 4.1. Alternatively, frequency bands are computed in small equal-sized bins, e.g. $\Delta f = 1\,\text{Hz}$ [127], or $\Delta f = 2\,\text{Hz}$ [109], [128], [129].

The mostly used algorithm to compute Discrete Fourier Transform is the Fast Fourier Transform (FFT) applied by [105], [109], [114], [130]–[136]. Commonly used alternatives are Short-Time Fourier Transform (STFT) [129], [137]–[139] or the estimation of Power Spectral Density (PSD) using Welch's method [116], [117], [119], [121], [127], [128], [140], [141]. When averaging over all windows, STFT is considered more robust against noise. Hence, this method is adopted in the experiment described in sections 4.4-4.6, using a Hamming window of length 1000 ms with no overlap as parameters. The features extracted from the resulting representation of the signal in frequency domain are: avrg. power (mean) of frequency bands and bins ($\Delta f = 2\,\text{Hz}$), their minimum, maximum, and variance. Additionally, the ratio of mean band powers β/α is calculated for each channel.

Higher Order Spectra (HOS)

The group of frequency domain features Also belonging to the group of frequency domain features are bispectra and bicoherence magnitudes used by Hosseini et al. [142], which are

---

[1]The backward difference $\nabla \equiv Z(t) - Z(t-1)$ is used here.

available in the *HOSA toolbox* [143]:

The bispectrum Bis represents the Fourier Transform of the third order moment of the signal, i.e.:

$$\text{Bis}(f_1, f_2) = \text{E}\left[\Xi(f_1) \cdot \Xi(f_2) \cdot \Xi^*(f_1 + f_2)\right], \tag{4.5}$$

where $\Xi(f)$ is the Fourier Transform of the signal $\xi(t)$, * denotes its complex conjugate, and $\text{E}[\cdot]$ stands for the expectation operation.

Bicoherence Bic is simply the normalized bispectrum:

$$\text{Bic}(f_1, f_2) = \frac{\text{Bis}(f_1, f_2)}{\sqrt{P(f_1) \cdot P(f_2) \cdot P(f_1 + f_2)}}, \tag{4.6}$$

where $P(f) = \text{E}\left[\Xi(f)\Xi^*(f)\right]$ is the power spectrum. Four different features are extracted for each electrode and each of 16 combinations of frequency bands, namely the sum and the sum of squares of both Bis and Bic (for details see [142]).

4.2.3 Time-Frequency domain features

If the signal is non-stationary, time-frequency methods can bring up additional information by considering dynamical changes.

Hilbert-Huang Spectrum (HHS)

Hadjidimitriou et al. compared three such methods, namely STFT based spectrogram (SPG, also used in [123]), the Zhao-Atlas-Marks (ZAM) distribution, and the Hilbert-Huang Spectrum (HHS) [110]. Although results were comparable between all methods, the latter, non-linear method, showed to be more resistant against noise than SPG. We hence computed HHS for each signal, which is done via empirical mode decomposition (EMD) to arrive at intrinsic mode functions (IMFs) to represent the original signal:

$$\xi(t) = \sum_{i=1}^{K} \text{IMF}_i(t) + e_K(t), \tag{4.7}$$

where $e_K(t)$ denotes the residue which is monotonic or constant. Using the Hilbert transform of each IMF_i, the analytical signal can be described by its amplitude $A_i(t)$ and instantaneous phase $\theta_i(t)$. The derivative of $\theta_i(t)$ can be used to compute the meaningful instantaneous frequency $f_i(t) = \frac{1}{2\pi}\frac{d\theta_i}{dt}$, which yields a time-frequency representation of the amplitude $A_i(t)$. Further, the squared amplitude is computed, and for both, the average of each frequency band is calculated as features.

Discrete Wavelet Transform

A more recent technique from signal processing is Discrete Wavelet Transform (DWT), which decomposes the signal in different approximation and detail levels corresponding to different frequency ranges, while conserving the time information of the signal. The trade-off herein is made by down-sampling the signal for each level. For details, the reader is referred to [144],

for example. Correspondence of frequency bands and wavelet decomposition levels depends on the sampling frequency and is given for our case of $f_s = 512\,\text{Hz}$ in the last column of Table 4.1 [145]. We implemented features from two wavelet functions as described below.

Murugappan et al. introduced feature extraction for emotion recognition from EEG via DWT using "db4" wavelet functions, extracting energy and entropy of 4th level detail coefficients (D4) corresponding to α band frequencies [113]. They extended their approach later to frequency bands β and γ, including root mean square (RMS), recursive energy efficiency (REE) and its log(REE) (LREE) and abs(log(REE)) (ALREE), as given in the equations below [118], [146]:

$$\text{RMS}(l) = \sqrt{\frac{\sum_{i=1}^{l}\sum_{j} D_{ij}^2}{\sum_{i=1}^{l} J_i}} \;, \tag{4.8}$$

where D_{ij} are the detail coefficients, J_i the number of coefficients at the i-th decomposition level, and l denotes the number of levels; and

$$\text{REE} = \frac{E_{\text{band}}}{E_{\text{total-3b}}} \;, \tag{4.9}$$

where E_{band} is the energy of a subband, and the total energy of subbands $E_{\text{total-3b}} = E_{\alpha} + E_{\beta} + E_{\gamma}$.

A different prototype wavelet was suggested by Frantzidis et al., i.e. third order biorthogonal wavelet functions [116]. Power of δ, β, and α bands are used as features.

4.2.4 Features calculated from combinations of electrodes

Multichannel complexity D2

Multi-channel D2 is introduced by Konstantinidis et al. as a non-linear measure representing the signal's complexity [62]. Unfortunately, description and citations in their paper are incomplete making it impossible to re-implement this feature[2].

Magnitude Squared Coherence Estimate (MSCE)

This feature (short MSCE) represents the correspondence of two signals ξ_i and ξ_j at each frequency, taking values between 0 and 1 [147]. It is defined as:

$$C_{ij}(f) = \frac{\left| P_{ij}(f) \right|^2}{P_i(f)P_j(f)} \;, \tag{4.10}$$

where P_{ij} is the cross power spectral density and P_i and P_j are the power spectral density of ξ_i and ξ_j, respectively. In order to reduce the large amount of features resulting from all possible combinations of electrodes, C_{ij} is averaged over the frequency bands. MSCE features are equivalent to cross-correlation coefficients used in other studies [105], [109], [121], [132], [148].

---

[2]Attempted communication with the authors could not resolve this issue, which is why no further attention is paid to this feature.
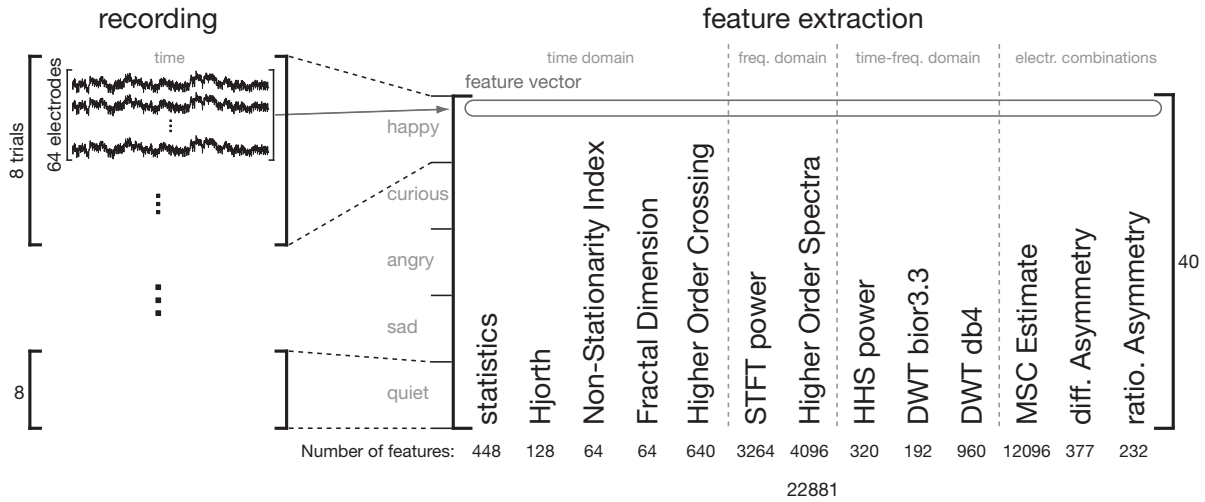
Figure 4.1: Schematic characterization of data processing from trial-wise recording to feature extraction.

Due to suggestions from neuro-scientific findings of hemispherical asymmetry related to emotion (e.g. [149]), some studies implemented features based on the combination of symmetrical pairs of electrodes. These features can be divided into differential asymmetry and rational asymmetry.

Differential asymmetry

Most frequently used are differences in power bands of corresponding pairs of electrodes, computed as the difference of two features:

$$\Delta x = x_l - x_r \ , \tag{4.11}$$

where (l, r) denote the symmetric pairs of electrodes on the left/right hemisphere of the scalp [127], [131], [136]–[141], [150]. Additionally, Liu and Sourina studied pairwise hemispherical differences of statistical features and Fractal Dimension [63].

Rational asymmetry

In few cases, ratios of features from symmetric electrodes are computed. Brown et al. used the spectrogram to compute α-power ratios x_l/x_r for consecutive time windows of $\Delta t = 2s$ [128]. From this, they extracted kurtosis, maximum, and the number of peaks. The latter was defined by a threshold of $\text{mean}(x_l/x_r) + 2\sigma$, where σ is the standard deviation, and normalized to the length of the recording. In analogy to the differential asymmetry features, some studies used the ratio of band powers in their study [138], [139].

4.3 Feature selection

The large amount of possible features makes it necessary to reduce this space in order to avoid over-specification and to make feature computation feasible online. Techniques to

Tabular 4.2: Studies on emotion recognition from EEG listed with used features and method to extract them, as well as the number/position of electrodes and methods for feature selection, if applied.

Author	Year	EEG Features	Method/Description	Electrodes	Feature Selection
Musha et al. [105]	1997	Cross-correlation form θ (5-8 Hz), α (8-13 Hz), β (13-20 Hz)	FFT	Fp1, Fp2, F3, F4, T3, T4, P3, P4, O1, O2	-
Choppin [151]	2000	α, β, θ, asym. diff., HOS, MSCE	PSD	Fp1, Fp2, F3, F4, T3, T4, P3, P4, O1, O2	heuristically
Takahashi [112]	2004	δ, θ, α, β, μ_ξ, σ_ξ, δ_ξ, γ_ξ, δ_ξ/σ_ξ, γ_ξ/σ_ξ	n/a	3 elec. headband	-
Bos [136]	2006	α, β, α/β	FFT	F3, F4, Fpz	-
Ansari et al. [115]	2007	Hjorth features (Activity, Mobility, and Complexity), fractal dimension	Sevcik's method	CP5, CP6, F3, F4, Afz	synchronization likelihood
Chanel et al. [129]	2007	9 bands 4-20 Hz ($\Delta f = 2$ Hz)	STFT	64 electrodes (IS)	FCBF
Murugappan et al. [113]	2007	entropy, energy of 4th level detail coeffs (α)	DWT (db4)	63/24 electrodes (IS)	24 electrodes: heuristically
Horlings [121]	2008	ERD/ERS, Cross-corr., Hjorth features	PSD (Welch's method)	32 electrodes (IS)	mRMR
Khalili and Moradi [152]	2008	θ (4-8 Hz), α (10-12 Hz), β, γ, δ_ξ	FFT	64 electrodes (IS)	GA wrapper
Schaaff and Schultz [109]	2009	AI: 34 bins of freq. components (5-40 Hz)	FFT	Fp1, Fp2, F7, F8	correlation coefficient
		AII: α, cross-correlation btw. electrodes different γ bands			
Li and Lu [135]	2009	different γ bands	FFT	62 electrodes (IS)	CSP, SVM wrapper
Petrantonakis and Hadjileontiadis [111], [126]	2010	Higher Order Crossings (HOC), DWT μ_ξ, σ_ξ, δ_ξ, γ_ξ, δ_ξ/σ_ξ, γ_ξ/σ_ξ	hybrid (adaptive) filtering (EMD & GA)	Fp1, Fp2, F3/F4	-
Khosrowabadi et al. [147]	2010	Magnitude Squares Coherence Estimate (1-35 Hz)	elliptic band-pass filter	P3, P4, T7, T8, C3, C4, F3, F4	-
Khosrowabadi and Rahman [123]	2010	TF spectogram with GMM (2-30 Hz), FD	STFT, Higushi/Box-counting/Brownian	P3, P4, T7, T8, C3, C4, F3, F4	univariate ranking

unless noted differently in parenthesis, frequency bands are defined as follows: δ: 1-4 Hz, θ: 4-8 Hz, slow α: 8-10 Hz, α: 8-12 Hz, β: 12-30 Hz, γ: 30-64 Hz

Tabular 4.2: continued.

Author	Year	EEG Features	Method/Description	Electrodes	Feature Selection
Lin et al. [139]	2010	δ, θ, α(8-13 Hz), β(13-30 Hz), γ(31-50 Hz) differential asym., rational asym.	STFT	30 electrodes (IS), 12 symmetric electr.	F-score index
Murugappan et al. [118], [146]	2010	energy, RMS, REE, LREE, ALREE, power, σ_ξ	DWT (db4)	64 electrodes (IS)	-
Hosseini et al. [142]	2010	bispectra and bicoherence magnitudes (Higher Order Spectra)	HOSA toolbox	Fp1, Fp2, T3, T4, Pz	GA wrapper (SVM)
Frantzidis et al. [116]	2010	δ(.5-4 Hz), θ, α(8-16 Hz) Ampl. and latency of ERPs	DWT (3rd order, biorth. wavelet)	Fz, Cz, Pz	SWM wrapper
Nie et al. [134]	2011	δ, θ, α(8-13 Hz), β(13-30 Hz), γ(36-40 Hz)	FFT, log band energy, feat. smoothing	62 electrodes (IS)	correlation coefficient
Kroupi et al. [119]	2011	θ, α(8-13 Hz), β(14-29 Hz), γ(30-47 Hz), normalized length density (NLD), non-stationarity index (NSI)	PSD (Welch's method)	32 electrodes (IS)	Fisher's meta analysis
Wang et al. [114]	2011	δ, θ, α, β, γ μ_ξ, σ_ξ, δ_ξ, γ_ξ, δ_ξ/σ_ξ, γ_ξ/σ_ξ	FFT, Hanning window	128 electrodes (IS)	mRMR
Makeig et al. [153]	2011	power of 8-200 Hz	FFT	128 electrodes (IS)	CSP
Brown et al. [128]	2011	max, kurtosis, peaks of asym. $\alpha(t)$ ratio avrg. power 6-12 Hz ($\Delta f = 2$ Hz)	spectrogram ($\Delta t = 2$s) PSD (Welch's method)	Fp1, Fp2, F3, F4, F7, F8, C3, C4	t-test, SFS
Soleymani et al. [141]	2011	θ, slow α, α, β, γ, differential freq. band asymmetries	PSD (Welch's method)	32 electrodes (IS), 14 symmetric electr.	(lin. ridge reg.)
Sourina and Liu [124]	2011	Fractal Dimension (FD)	Higuchi, Box-counting	Af3, F4, Fc6	-
Liu and Sourina [133]	2012	β/α ratio, β	FFT	Af3, F7, F3, Fc5, Fc6, F4, F8, Af4, P7, P8	-
Konstantinidis et al. [62]	2012	multichannel complexity D2	Correlation Integral	19 electrodes (IS)	-

unless noted differently in parenthesis, frequency bands are defined as follows: δ: 1-4 Hz, θ: 4-8 Hz, slow α: 8-10 Hz, α: 8-12 Hz, β: 12-30 Hz, γ: 30-64 Hz

Tabular 4.2: continued.

Author	Year	EEG Features	Method/Description	Electrodes	Feature Selection
Hadjidimitriou and Hadjileontiadis [110]	2012	TF analysis of β (13-30 Hz) and γ (30-49 Hz) with rest norm.	spectrogram, ZAM, HHS	14 electrodes (IS)	-
Rozgić et al. [140]	2013	θ, slow α, α, β, γ, differential asymmetries	FFT	32 electrodes (IS)	(seg.-level feat. transform)
Reuderink et al. [127]	2013	1-64 Hz ($\Delta f = 1$ Hz), diff. asym. of α	PSD (Welch's method)	32 electrodes (IS)	correlation analysis
Liu and Sourina [63]	2013	diff. asym. of statist., FD, δ, θ, α, β, γ	n.a.	32/14 electrodes (IS)	univ. SVM wrapper
Park et al. [150]	2013	θ, α, β, γ, MSCE, diff. asym.	n.a.	Cz, P3, P4, F3, F4	-
Duan et al. [138]	2013	δ, θ, α (8-13 Hz), β (14-30 Hz), γ (31-50 Hz) differential asym., rational asym.	STFT, diff. entropy	62 electrodes (IS) 27 symmetric electr.	mRMR
Bahari and Janghorbani [154]	2013	phase space features from Recurrence Quantification Analysis		32 electrodes (IS)	t-test, ROC
Lee and Hsieh [132]	2014	cross-corr., MSCE, phase sync. (θ, α, β, γ)	FFT	64 electrodes (IS)	correlation
Jie et al. [155]	2014	sample entropy (β)	DWT	32 electrodes (IS)	K-S test
Lan et al. [117]	2014	θ, α, β, FD, HOC μ_ξ, σ_ξ, δ_ξ, γ_ξ, δ_ξ/σ_ξ, γ_ξ/σ_ξ	PSD, Higuchi	FC5, F4, F7, AF3, T7	intra-class correlation
Lee et al. [137]	2014	diff. asym. of α, γ	STFT	11 electrodes (IS)	-
Jirayucharoensak et al. [131]	2014	θ, slow α, fast α (10-12 Hz), β, γ, differential asymmetries	FFT, Hanning window	32 electrodes (IS)	(PCA)
Conneau and Essid [130]	2014	power, moments, shape descriptors (θ, α, β, γ)	FFT	54 electrodes (IS)	CSP
Liu et al. [156]	2014	θ, α, low β (13-20 Hz), high β (20-30 Hz) γ	band-pass filter	64 electrodes (IS)	(PCA, KFDA)
Vijayan et al. [148]	2015	Shannon entropy, cross-corr., AR coeff. (γ)	DWT (db5)	12 electrodes	-

unless noted differently in parenthesis, frequency bands are defined as follows: δ: 1-4 Hz, θ: 4-8 Hz, slow α: 8-10 Hz, α: 8-12 Hz, β: 12-30 Hz, γ: 30-64 Hz

achieve this have been used in some of the previous studies (see last column of Table 4.2), but they are limited to selecting from the set of features included in each study. Here, we aim to systematically compare features on *one* dataset to draw more general conclusions. As mentioned above, results obtained from applying multiple FS methods are considered more robust [82]. We apply five state-of-the-art filter methods which are introduced and examined w.r.t. their capabilities of detecting interacting features in Ch. 3. These are the univariate filter methods ReliefF [85] and ES f^2 [88] as well as the multivariate filter methods mRMR [91], ES Λ [97], and the newly proposed ES θ. As done earlier in Ch. 3, we chose 20 levels of discretization required for the FS method mRMR.

We guarantee the requirement that features are uncorrelated (see Sec. 3.2.2) by a general preprocessing step explained in Section 4.5. It can be noted that using a dataset with multiple classes as recorded in this work (see next section) actually benefits the expressiveness of the FS results, since it becomes less likely, compared to fewer classes, that a random set of features provides a good separability [78].

4.4 Database

To conduct an experiment comparing features listed in Section 4.2, we recorded a reasonably large database. Of special importance was to achieve a reasonably high spatial resolution of electrodes as well as to cover a broad range of emotions.

4.4.1 Experiment and procedure

In total, 16 subjects (7 female, 9 male) aging between 21 and 32 participated in the experiment. For each subject, the recorded dataset contains 8 trials of 30s EEG recording for 5 different emotions (happy, curious, angry, sad, quiet), which were selected to cover a good part of the valence-arousal-dominance (VAD) space. The experiment took place in a closed environment controlled for sound and light conditions. Ethical approval of the experiment was obtained from the ethics committee of the medical faculty of TUM. Emotions were induced using IAPS pictures [41]. We selected pictures on the basis of dimensional ratings, that are included in the IAPS database, by means of an in-house developed tool (see Appendix A.1) to find best matches around a user defined mean with minimal variance. Means of above mentioned emotions in VAD space were taken from Russell and Mehrabian [26][3]. This way, a total of 32 pictures were selected for each emotion, i.e. a total of 160 emotional pictures.

Subjects were instructed on the experiment procedure and filled out a questionnaire testing for suitability (e.g. skin allergies), well-being (WHO), and trauma screening (PTSD). No subject had to be excluded. After being equipped with recording sensors, a test-run using only neutral IAPS pictures was conducted prior to the actual experiment to accustom the subjects to the protocol and the concept of the SAM questionnaire (Self-Assessment-Manikin Test [45]), which was completed after each trial. Figure 4.2 shows the induction protocol implemented. Each trial started with a black screen, followed by 4 randomly drawn pictures

---

[3]Since emotions angry and disgust overlap in VAD space, some angry pictures were chosen by hand.
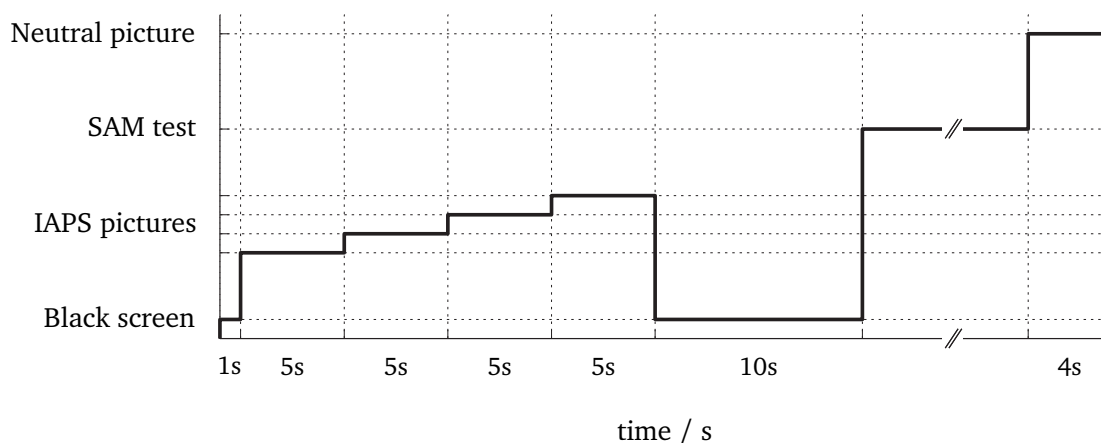
Figure 4.2: The emotion induction protocol for each trial consisting of the set of IAPS pictures, prolonged recording time, SAM test, and a neutral picture.

from an emotion subset, each shown for 5 seconds. Every picture was used only once per subject. To capture late effects of emotional responses, recording of each trial was prolonged by 10*s* showing a black screen before SAM ratings on a scale from [1; 9] were taken. Each trial ended with a neutral picture (4*s*) to reduce the effect of biased transition probability based on the previous emotion [157]. In total, 40 trials were recorded for each subject, while the order of emotions was randomized.

A 64-channel EEG cap with g.tec gUSBamp and electrodes placed according to the extended 10-20 system was used for recording at 512 Hz. The amplifier was set to include a 50 Hz notch filter and a band-pass filter ($0.1 - 100$ Hz). The effect of artifact removal, which is a common pre-processing step for EEG data, was found to be marginal and thus, was not carried out. This speeds up processing substantially, i.e. favors online recognition capability.

4.4.2 Evaluation of induction

Since successful induction of emotion is a major challenge, we tested whether the average of dimensional rating over each combination of four pictures used for induction corresponds to the targeted emotion range. Although the IAPS pictures do not resemble the emotion values to the same extreme as literature suggests, all combinations overlapped with the target emotion region and were well separable. This particularly holds for the hand selected angry pictures.

For validation of induction, we compared the results of SAM tests with the targeted values of the presented picture sets and received an average correlation coefficient of $r = .545$. The second column in Table 4.3 lists the correlation coefficient for each subject. Its variance is very large with the extreme values of .255 and .848, which suggests a highly individual reaction to emotion stimuli. Particularly high correlation and supposedly successful induction of emotion was reached for subjects 1, 7, 10, 11, and 13. We further discuss our results with respect to the induction success in Section 4.6.

4.4.3 Evaluation of EEG data

Albeit careful preparation and signal checks previous to the experiment, its duration involved the possibility to loose proper contact of shifting electrodes. We therefore analyzed the signal quality of each subject. While for some subjects few electrode signals were permanently noisy over the whole duration of the experiment, for others signal problems occurred in some trials for almost all electrodes. Since these missing trials would reduce the already limited amount of samples per subject, we decided to exclude these subjects from further processing. As depicted in the third column of Table 4.3 noise corrupted electrodes were found for subject 1, 5, 7, 11, 16.

4.5 Classification

Pre-studies showed that it takes about 10 s for an emotion to be induced using pictures. The duration of emotions in laboratory environments is typically assumed $0.5-4$ seconds [158]. Thus, time intervals between 11-15 s after emotion induction onset, i.e. the appearance of the first picture, were considered in the following analysis. Given a sampling rate of 512 Hz, this resulted in 2048 time samples in our case.

Figure 4.1 illustrates data processing for each subject. A feature matrix was generated from the EEG data of 8 trials and 5 classes, leading to 40 rows or feature vectors. Features were extracted from all 64 electrodes which results in a total of 22881 features. The resulting number of features is indicated for each feature group. Features like MSCE, which was computed from all combinations of electrodes and frequency-ranges resulted in 12096 features alone. Except for MSCE and asymmetry features, all features can be associated with one electrode, which allows us to draw conclusions on their importance for emotion recognition in Section 4.6.3. Features were z-normalized to zero mean and standard deviation equal to one. Further, we eliminated all almost identical features which yield a correlation coefficient higher than .98 in order to avoid problems with singularities, which may occur for some FS methods.

Given the relatively small number of samples for each class, leave-one-out cross validation was chosen. We evaluated each of the five listed FS methods for each subject individually. To avoid overfitting and to limit the computational effort, the maximum number of features was fixed to 200. The data were divided into 8 folds with 5 samples each for cross-validation. Thereof, 7 folds were used for feature selection and training of the classifier. In turn, the remaining fold (i.e. 1 sample from each class) was used for testing, until each fold was tested once. Given 5 classes, chance level was at 20%.

To evaluate and compare the proposed feature selection methods, classification was performed by means of quadratic discriminant analysis (QDA) with diagonal covariance estimates (i.e. Naive Bayes). We chose this classifier, since it is generally known as a robust baseline classifier.

4.6 Results

As seen in Section 4.4 and as discussed within the community of affective computing, reliable induction remains an ongoing challenge in the field. However, given the recorded dataset and experimental results, we believe to be able to contribute some insights to the following three questions, which we discuss in the following subsections: 1) How do the different feature selection techniques perform and how many features are generally required? 2) Which type of features are most successful? Can we draw conclusions w.r.t. which features are generally better across subjects? 3) Which electrodes are mostly selected? Are there commonalities/differences regarding electrode selection across feature types?

4.6.1 FS methods

Table 4.3 lists achieved accuracies at the optimal number of features in tabular form over subjects and FS methods. The best result for each subject is highlighted in bold font. On average, multivariate methods (mRMR, ES Λ, ES θ) perform slightly better than univariate methods (ReliefF, ES f^2), as can be seen in the bottom row. In particular, the average optimal number of features using ES Λ, i.e. 44.6, is small. Less than 100 features per subject are chosen on average over all FS methods (see last column in Table 4.3). However, the optimal number of features varies strongly between subjects. It should also be noted that stable results, i.e. accuracy does not vary anymore starting at around 30 features for most subjects and methods.
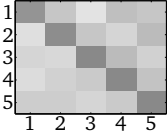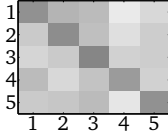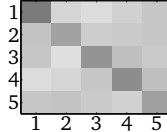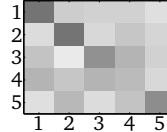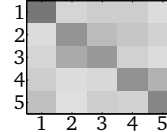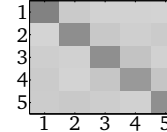
Classification accuracy ranges subject-dependent from 25.0% to 47.5%, which correlates with the evaluation of induction, i.e. the SAM test results ($r = .56$). In this, however, FS using ReliefF shows almost no correlation (.05), while using ES Λ yields $r = .61$.

Confusion Matrices (presented in Table 4.3 and Fig. 4.3) summarize class-dependent results across subjects and methods. Noticeably, FS using ES Λ reveals problems recognizing emotion class *sad*, while *happy* and *curious* are mostly identified correctly. Feature selection using ES θ is more balanced across classes, i.e. the diagonal values are high compared to non-diagonal values. But with this method (and also ReliefF) emotions *curious* and *angry* are sometimes confused with each other. Most stable and evenly distributed results are obtained from feature selection using mRMR. Overall, emotion *happy* is generally better recognized than emotion *sad*. Between subjects, we see a high variance of emotions well recognized. Subjects 3, 12, and 13, for example, show good separability of only one or two emotions from the rest (see darker squares in confusion matrices). A more even separability of classes is given for subjects 4, 6, 10, and 15. Emotion *quiet* shows smallest variance in recognition rates across subjects.

4.6.2 Feature usage

To investigate which features are superior to others, i.e. have been selected more frequently by FS methods and performed more successfully in emotion classification, we compute the relative frequency of each of the 162 feature types. In this, we create a histogram of feature occurrence given the features selected at the optimal number of features for each subject and FS method. Then, we normalize each bin of the histogram by dividing the occurrence

Tabular 4.3: Pair-wise correlation coefficients between IAPS and SAM ratings, signal quality and accuracy (optimal number of features) for different FS methods listed for each subject.

Subject	Correlation coefficient	Signal Quality	mRMR	ReliefF	ES f^2	ES Λ	ES θ	Average
1	.748	✗	-	-	-	-	-	-
2	.588	✓	35.0% (84)	**37.5%** (56)	32.5% (66)	**37.5%** (25)	32.5% (131)	35.0% (72.4)
3	.524	✓	27.5% (196)	37.5% (17)	**42.5%** (4)	32.5% (5)	35.0% (7)	35.0% (45.8)
4	.711	✓	**35.0%** (82)	**35.0%** (29)	30.0% (10)	**35.0%** (31)	27.5% (4)	32.5% (31.2)
5	.447	✗	-	-	-	-	-	-
6	.589	✓	**45.0%** (30)	37.5% (83)	37.5% (31)	35.0% (23)	35.0% (3)	38.0% (34.0)
7	.848	✗	-	-	-	-	-	-
8	.573	✓	35.0% (122)	27.5% (12)	**40.0%** (49)	35.0% (138)	37.5% (37)	35.0% (71.6)
9	.490	✓	32.5% (19)	**37.5%** (132)	30.0% (149)	30.0% (74)	30.0% (78)	32.0% (90.4)
10	.748	✓	35.0% (192)	42.5% (75)	32.5% (121)	**45.0%** (92)	40.0% (18)	39.0% (99.6)
11	.794	✗	-	-	-	-	-	-
12	.511	✓	37.5% (70)	32.5% (49)	25.0% (149)	37.5% (6)	**42.5%** (195)	35.0% (93.8)
13	.796	✓	**45.0%** (4)	37.5% (3)	**45.0%** (183)	42.5% (40)	**45.0%** (151)	43.0% (76.2)
14	.255	✓	35.0% (10)	**40.0%** (40)	30.0% (113)	35.0% (16)	32.5% (180)	34.5% (71.8)
15	.550	✓	40.0% (106)	25.0% (67)	35.0% (133)	35.0% (35)	**47.5%** (109)	36.5% (90.0)
16	.295	✗	-	-	-	-	-	-
		Average:	36.5% (87.1)	35.4% (68.8)	34.5% (82.1)	36.3% (44.6)	36.8% (82.6)	35.9% (70.6)

Legend: 1 - happy
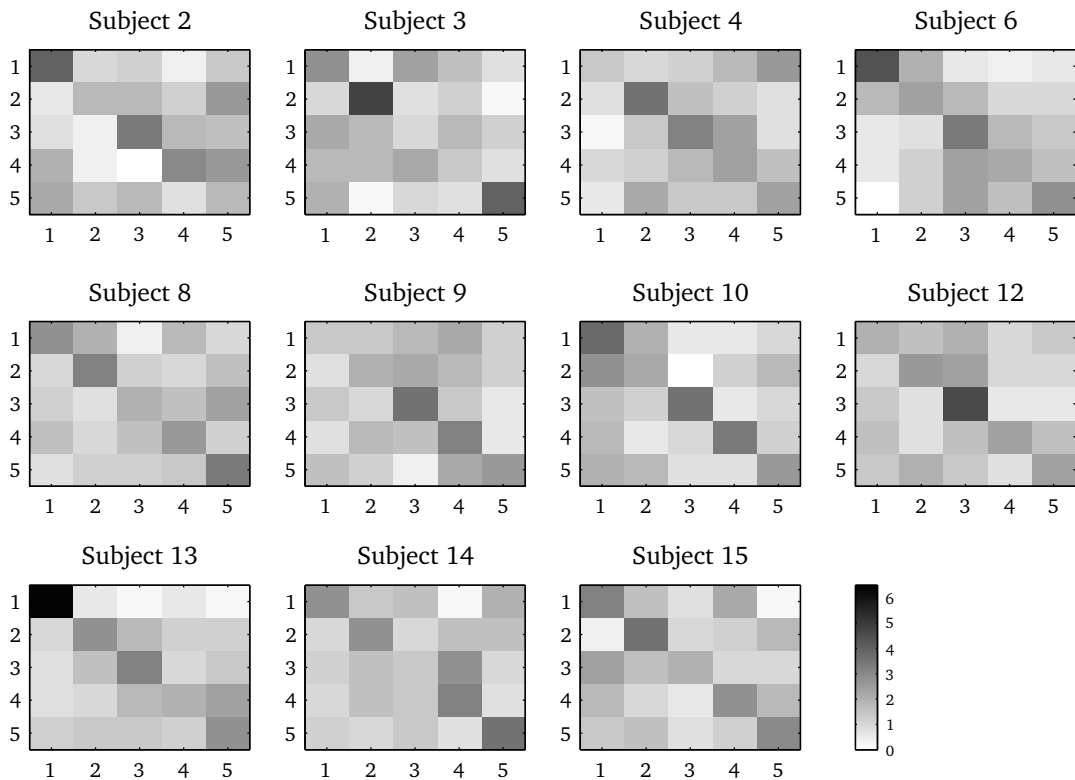2 - curios
3 - angry
4 - sad
5 - quiet

Figure 4.3: Confusion matrices summarizing the targeted (y-axis) and predicted (x-axis) emotion class are depicted for each subject individually.

of a feature type by the number of features belonging to this feature type (e.g. 64 in the case of feature type max α band powers, i.e. one feature for every channel) to account for random assignment of features (this is important since combinations of features can not be associated with a single electrode and, thus, are each considered one type of features). We further weighted these relative frequencies to account for the following: 1) when averaging across FS methods, multiply each relative frequency with the accuracy achieved to account for different success rates of FS methods, and 2) when averaging across subjects, multiply each subject's relative frequency with the SAM correlation coefficient to account for general reliability of the data. Given this statistic, we assume that valuable features score higher than features not important for successful emotion recognition. Fig. 4.4 illustrates the weighted relative occurrence of each feature type together with the average and standard deviation of subgroups of features. In general, commonalities/differences between subjects of different induction and accuracy level are not very apparent. The tendencies visible from the data are discussed in the following:

On the one hand, the most frequently selected feature group is HOC features, including a tendency to favor features with large values for k. Also, features measuring complexity of the time sequence, i.e. fractal dimension, NSI, and Hjorth's Complexity, score well. The most valuable feature type among the combinations of electrodes are the rational asymmetry features. Other (sub)-groups, such as HHS and HOS bicoherence features, yield high mean values on the weighted relative occurrence, too, but show a higher variance than the HOC
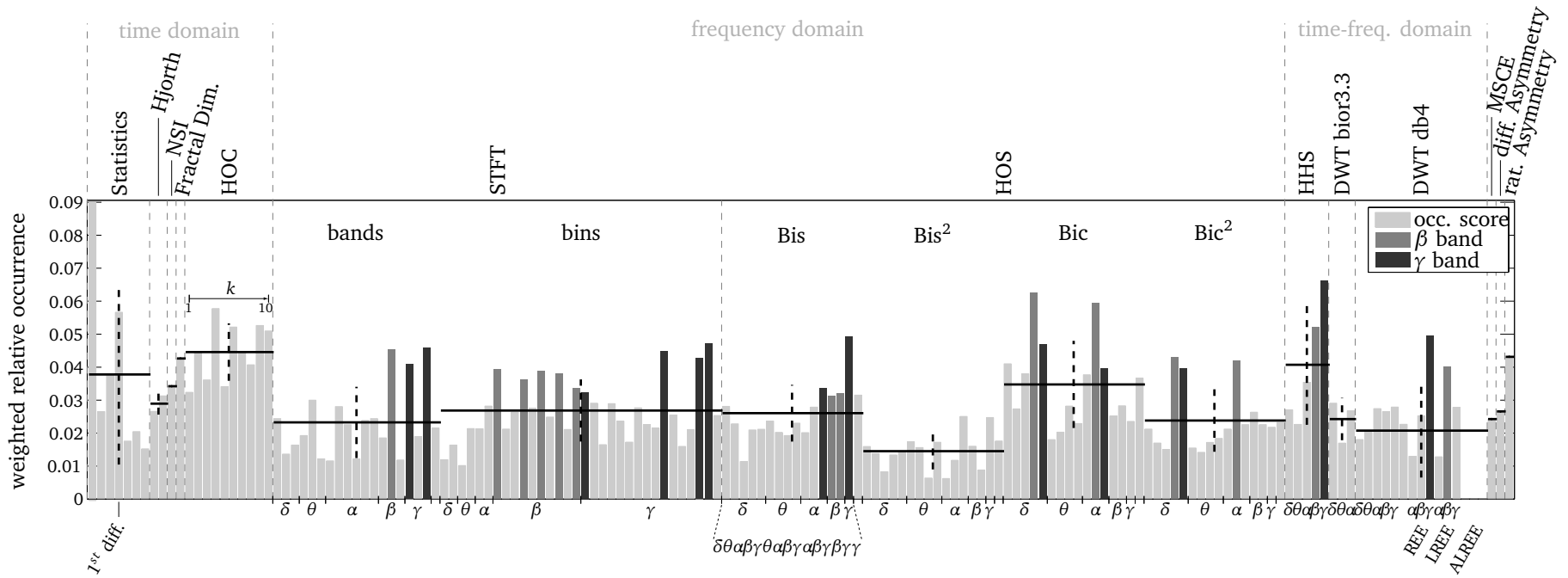
Figure 4.4: Feature type usage is measured by weighted relative occurrence scores (gray bars). Mean (solid horizontal lines) and variance (dashed vertical lines) are shown for each subgroup of feature types. Feature types computed from certain frequency bands as well as individual feature types discussed in Sec. 4.6.2 are indicated below the figure.

group. This means that particular feature types of this group are more valuable than others.

Taking a closer look at the feature types reveals that especially features computed from frequency bands β and γ are successfully selected more often than other bands (bars marked darker in Fig. 4.4). This holds for feature groups STFT, HOS, HHS and DWT db4 features REE and LREE.

On the other hand, results suggest that feature subgroups such as HOS squared bispectra and DWT db4 are less efficient for emotion recognition from EEG, since their weighted relative occurrence scores are low. Feature ALREE apparently has no advantage over LREE, since it has never been selected. Both the prevalent frequency band powers and equal sized bins show relatively low scores on average.

The group of statistical features affords a differentiation between feature types, since the proposed statistics are very different from each other. Except for the first difference of the time series, none reveals a very high score on the weighted relative frequencies. A special note regarding the high score of the statistical feature power of signal should be given: This does not mean to be a highly valued feature, but has to be explained by a very high occurrence when using FS method ES θ: the first feature is selected in case of equal ranking scores or no measurable improvement through any feature (saturation of selection criterion).

4.6.3 Electrode usage

Based on the findings of Section 4.6.2, we investigated also electrode usage of feature types and groups that can be associated with a single electrode, i.e. all except MSCE and asymmetry features. In this section, we discuss the electrode usage of the six features identified as most important. These are 1st difference, complexity features, the group of HOC features and bicoherence features, and selected (threshold at .03) β, and γ features (i.e. from STFT, bispectrum, absolute and squared bicoherence, and REE(γ)/LREE(β)) as shown in Fig. 4.5. Here, the darkness of each electrode depicts the number of its occurrences within a feature group, where the scale is normalized by the total number of feature occurrences in the group, so that the scores of all 64 electrodes sum to one (since the first plot only captures electrode selections of one feature type, i.e. 64 features, we expect the results to be more pronounced here).

For the feature type 1st difference δ_ξ the prominent electrodes are P9 and P5 with some importance given to P7, P1, and FC2. The three complexity features Hjorth's complexity, NSI, and fractal dimension are mostly drawn upon from electrode locations CP2, CP1, P1, and F5. For the feature group of HOC the important electrodes are CP1 and CP2, but also some right hemisphere locations like T8 and P10 are frequently used. Features for bicoherence are mostly selected from C2, FC2, and Cz. A cluster of electrodes selected for good γ features forms around CP5, CP3, CP1, P5, and P3. Additionally, location CP6 is selected often. Not quite as clear are location of good β features. Prominent electrodes are C2, Cz, P9, F7, CP5, and CP3. Generally, electrodes located over the parietal and central-parietal lobe are favored over occipital and frontal lobes.
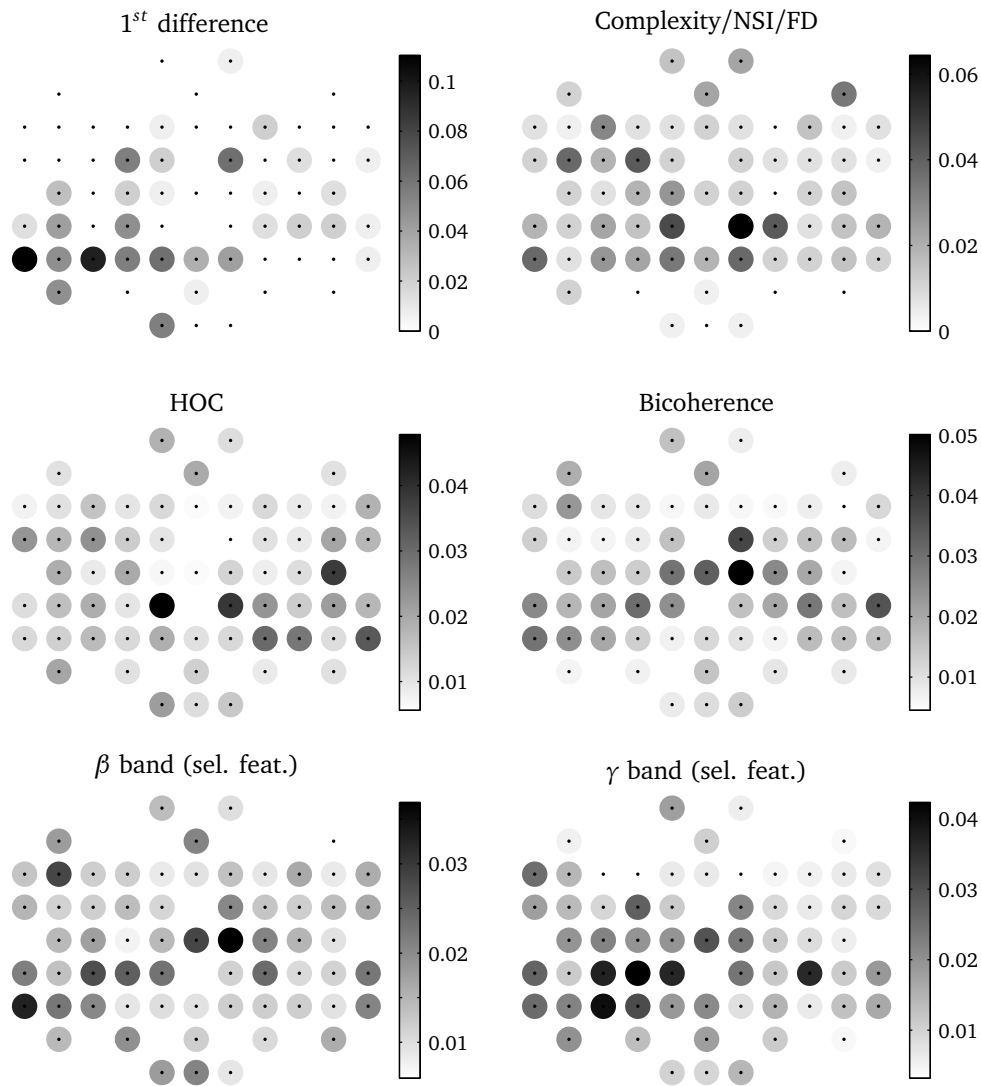
Figure 4.5: Electrode usage of six feature (sub-)groups is indicated by the darkness of the circle around each electrode location. Rows correspond to letters, columns to numbers of the extended 10-20 system.

4.7 Discussion

In this section, we interpret and discuss the same underlying questions posed above using results presented in Section 4.6.

The higher average accuracy of multivariate methods points towards interactions of features that only multivariate FS methods can actively detect and capitalize upon. The insights gained from Ch. 3 allow for some speculations on the underlying interactions, but have to be put in perspective given the relatively small number of samples per class. For example, since ES Λ and ES θ achieve comparable accuracies, it could be supposed that less complicated interactions like intra-class covariance between features are present. In combination with the good results of mRMR, this could be interpreted as evidence for negative interactions. High

variance in measured induction success as well as in emotion recognition accuracy between subjects, however, reassures the need for ongoing efforts towards a more reliable ground truth.

Spectral power in nearly all frequency bands is used in most studies, which contrasts with its low performance scores in our study. Among these features, Wang et al. found frontal and parietal locations typical under their TOP-30 features [114]. However, features related to β and γ bands seem more valuable if obtained by different (mostly more complex) extraction methods which have found only little attention so far. Moreover, our study revealed that electrode locations over parietal and centro-parietal lobes can be of advantage for these bands. Although the difference between the two STFT feature subgroups is small, the trend of equal sized bins scoring higher than frequency bands is in accordance with earlier argumentation by Reuderink et al. [127], that emotional responses can be visible in small frequency changes. The finding of β and γ bands from HHS being valuable features is in line with findings of Hadjidimitriou and Hadjileontiadis, who found these bands to discriminate best between liking and disliking judgments [110]. The usage of several complexity measures (especially FD) by many of the existing studies is endorsed by our results. The suggested electrode locations over the left anterior scalp as well as central-posterior locations are in line with those indicated by Ansari et al. [115]. A promising, yet little considered feature is HOC. Contrary to the frontal electrode used by Petrantonakis and Hadjileontiadis [111], centro-parietal locations might enhance this feature.

To date, frontal regions are of high importance. They are used in almost every study, especially when only a few electrodes are hand-selected. On the one hand, our results opposing this trend might be biased since we did not use artifact removal in this study. On the other hand, recent studies reach consensus that frontal alpha asymmetry primarily reflects approach vs. withdrawal motivation rather than the level of valence [159]. This may imply that locations other than the anterior scalp are better to differentiate multiple, discrete emotional states. Finally, rational asymmetry seems to have advantages over differential asymmetry, although the latter is used more frequently in the surveyed studies.

When interpreting these results it should be noted that due to the high number of tested features compared to the low number of samples, we could not validate statistical significance of our results.

4.8 Conclusion

Different sets of features and electrodes for emotion recognition from EEG have been suggested by over 30 studies reviewed in this chapter. We presented a systematic analysis and first qualitative insights comparing the wide range of available feature extraction methods using machine learning techniques for feature selection. Multivariate feature selection techniques performed slightly better than univariate methods, generally requiring less than 100 features on average. We also investigated which type of features are most promising and which electrodes are mostly selected for them. Advanced feature extraction methods such as HOC, HOS, and HHS were found to outperform commonly used spectral power bands. Results suggest preference to locations over parietal and centro-parietal lobes.

As shown by Mühl et al., different induction modalities result in modality-specific EEG

responses [104]. This possibly limits our results to the induction modality of visual stimuli. Due to the constant proposal of new features, it might be useful to enlarge the list of features from time to time. Yet, our presented methodology of comparing feature systematically remains valid. We expect that repeated experiments of the procedure on other databases will help to reach a consensus on these topics.

Another compelling question to take a closer look at in the future is whether certain types of features work better together than by themselves.

Part II

Dynamic emotion recognition

5 A dynamic model incorporating appraisal into emotion recognition

Summary. *Recent work in emotion theory points out the dynamic nature of emotion processes. However, the dynamic evolvement of the affective state has hitherto been neglected in the field of emotion recognition. Thus, a reconsideration of the approach to designing systems for emotion recognition might be fruitful and should take into account knowledge from the theoretical domain. In doing so, the main contributions of this chapter are*

- *a dynamic framework that is capable of incorporating appraisal theory into emotion recognition,*

- *a model combining findings and assumptions from modern emotion theory with data-driven methods to identify parameters,*

- *an in-detail version of the model for estimating emotion intensity in a specific experimental design.*

To date, most work on emotion recognition concerns itself with static prediction of emotion labels from a window of time series data using machine learning methods. Despite the shift of affective research towards capturing more subtle affective states, e.g. by using dimensional emotion models, the reliability of ground truth, duration and intensity of emotion are pertinent issues that have hardly been addressed [44]. Although reliability of induction is often identified as a limiting factor, most studies are relying on operator labeled blocks of recordings to predict a targeted induced emotion label. In recent studies, this issue has sometimes been tackled by trying to identify phases of "strong feeling" post hoc, but intensity of emotion and particularly its evolvement over time is mostly left unregarded [160]. This is probably also due to the fact that experiments that capture dynamics are obviously harder to design.

5.1 Problem statement

Static emotion recognition and machine learning approaches were important to understand basic differences in emotions and their general interaction with and effect on physiological signals. The ultimate goal of this field being sophisticated applications in real-world settings, however, makes an understanding of the whole emotion process necessary. Thus, it might be fruitful to consider findings from modern emotion theory such as appraisal models. Mortillaro et al. has advocated to use appraisal in emotion recognition by suggesting to use appraisal as an additional layer in between expressive features and the predicted emotion

label [34]. Yet, the most prominent finding is the fact that emotions are of dynamic nature and therefore, "require a dynamic computational architecture" [32]. We endorse this view and suggest that even the dynamic evolvement of the affective state should be considered in emotion recognition.

The main research objective in this chapter is to design a gray-box model for emotion recognition from physiological signals, which is capable of incorporating findings and assumptions from appraisal models, specifically Scherer's Component Process Model (CPM), while identifying parameters using data-driven methods from experimental recordings. Two main contributions are made within this chapter: 1) We review the literature on the intersection of emotion dynamics and emotion recognition to identify possible ways for the integration of appraisal into emotion recognition. 2) We propose a dynamic framework for intensity estimation based on the Dynamic Field Theory (DFT) and discuss an in-detail version of the model for a specific experimental design. The subsequent Chapter 6 reports an exemplary study implementing and evaluating this model.

The remainder of this chapter is organized as follows: Section 5.2 presents the state of the art on related topics. Our general approach to combine appraisal with data-driven methods is presented in Section 5.3. Section 5.4 presents the dynamic model and discusses its main properties. We conclude this chapter with some general remarks on the model in Section 5.5.

5.2 State of the art

In order to lay out a common ground for our approach, different fields of interest related to the problem stated above are reviewed in this section: 1) the representation of the affective state, 2) investigative work on emotion dynamics as such, 3) predictive approaches from physiological signals as well as relevant work from other modalities, and 4) dynamic measurement of affective labels.

5.2.1 Representation of emotions

As we debated in Section 2.1, appraisal models take on a more general view trying to consider and model the whole emotion process instead of just its representation. In particular for the CPM, this includes the five involved functional components: the appraisal processes, the motivational component, the physiological efferent effects in the automatic nervous system (ANS), the motor expression component in the somatic nervous system (SNS), and the subjective feeling component. Figure 5.1 shows how each component interacts with each other and, in particular, the subjective feeling, which we are concerned with in emotion recognition. In this part of the thesis, we adopt the terms used by the CPM, i.e. emotion quality, intensity, and duration, to characterize an affective state.

5.2.2 Emotion dynamics

To a large extend, dynamics of emotions have been neglected in emotion research, especially in the field of emotion recognition. But even in psychological modeling of emotions, dynamics have not played a big role for a long time. In contrast to most other emotion models,

appraisal models explicitly consider the temporal unfolding of emotion processes [32]. Dynamics of emotions, however, can be identified on various levels: Firstly, the chronological order of processes happening (type 1), e.g. appraisal checks, is considered an important research question. Sander et al. revises some studies which investigated this issue for Scherer's CPM [161]. Secondly, different durations of affective states (type 2) are termed in different ways (e.g., emotions, moods, and traits [20]). A third form of emotion dynamics is found in transitions between two or more emotion qualities (type 3). The influence of a previous affective state on the current affective state has been modeled by Karg et al. using piecewise linear systems [157], [162]. Another study on transitions between shape models of facial expressions of emotions in activation-evaluation space was done by Hakim et al. [163]. They found emotion trajectories in two-dimensional space to be smooth and persistent with time and that the paths traveled depend on whether starting and ending emotions correlated positively or negatively. Last, changes in emotion intensity over time (type 4) account for the dynamic unfolding of one emotion quality. The question of the dynamics of this highly complex process, and directly related its ramp-up and decay rates, does not yet have a satisfying answer [164]. We will refer to these different aspects of emotion dynamics as type 1-4 throughout this part of the thesis.

There are only a handful of studies that investigated intensity of emotion together with physiological signals. Statements on the dynamic nature, however, are rare. Waugh et al. studied models for BOLD response of discrete emotional intensity levels [165]. He found time-varying models to fit the data better than models conventionally applied to model BOLD signals. Other neuro-scientific studies based on fMRI have investigated intensity regarding involved brain regions and people's ability to recognize intensity of emotion from facial expressions [166]–[168].

A real world study on stress levels from Galvanic Skin Response (GSR) data was carried out by Bakker et al. [169]. The difficulty of investigating dynamics in combination with lack of control over the experiment, i.e. on the different kinds of events that influence physiological signals, limited their study to compare two statistical methods for detection of change in the signal, while the link to stress levels could not be further investigated.

Dynamic aspects of the subjective feeling component were shown by Grewe et al., who studied the influence of music on changes in both subjective feeling and physiological signals over time [40]. Although they found evidence against the possibility of inducing emotions using music, they found a number of events where skin conductivity changes at some novel events (e.g. choir entry in music piece) in correlation with subjective feeling ratings, but not at other similar non-novel events. This is interpreted as part of an orienting reflex, which is also part of the stimulus evaluation checks (SECs). Ratings were recorded in VA space using *EMuJoy* (see section 5.2.4). The authors acknowledge the serious problem of multi-dimensional self-report, demanding subjects to perform a *mental principle component analysis* online to accurately assess their affective state.

5.2.3 Predictive approaches

Out of the large body of literature on emotion recognition from physiological signals, only a few fulfill one of the following criteria related to the work in this chapter: 1) consideration

of dynamics and 2) estimation of emotion intensity. Additionally, predictive studies that consider 1) and 2) from other modalities are listed.

Emotion recognition from physiological signals mostly concerns itself with static prediction of emotion quality, i.e. estimates a discrete affective state given a certain window of time series data, but neglects the consideration of emotion dynamics. This is not to be confused with features that capture dynamic properties of a signal within a certain time window that is being evaluated for its emotional content. For overviews of features from both peripheral physiological features and EEG features see [51] and [201], respectively. For example, Valenza et al. used a quadratic nonlinear autoregressive integrative (NIRA) model to extract linear and nonlinear features from short-time electrocardiogram (ECG) [170]. Kulic and Croft used reactions from human-robot-arm-interaction in peripheral physiological signals to study emotion classification of three discrete levels of arousal and valence using Hidden Markov Models (HMMs) to represent time-domain observation sequences [171]. Yet, these traditional machine learning problems do not consider the dynamic evolution of the affective state itself.

Prediction of intensity on a continuous scale from physiological data has only been published by Bailenson et al. [172], to the best of our knowledge. They performed regression on both facial and physiological features to estimate intensity levels of the emotion qualities amusement and sadness every second. In this approach, the estimation of quality and intensity are separated, which we adopt similarly. Features from physiological signals were found to improve evaluation measures, compared to facial features only, especially for intensity of sadness. Although this method constitutes a high rate of prediction, it does not represent a dynamic model of emotion.

Some relevant studies can be added when borrowing from other modalities as a source for prediction of emotion intensities. Kim et al. recently published a study on estimating the intensity of conflict on a continuous scale using various regression methods on audio features from debates [173]. Roether et al. extracted dynamic features of gait, e.g. speed, to predict continuous intensity labels of five emotions using sparse regression [174]. They found speed to correlate positively with intensity of emotions happy and angry, and negatively with intensity of emotions sad and fear. In the field of emotion recognition from speech, Wöllmer et al. introduced Long Short-Term Memory Recurrent Neural Networks to predict the quadrant in VA-space quasi-continuously [175]. This method allows memorizing information, thus implicitly taking dynamic properties of emotions into account. A truly dynamic approach to modeling emotional content of music using system identification is given by Korhonen et al. [176]. The capability of different state-space and ARX models to simulate VA labels continuously over time is successfully tested for one genre of music.

5.2.4 Dynamic measurement of affective states

Controlled induction of emotion is difficult, since reactions of subjects can vary to the same stimuli, even when presenting the same stimulus to the same subject. This is a major problem in current emotion recognition research making the evaluation of induction as applied in the experiments reported in Section 4.4.2 and Section 6.1.5 essential. Karg et al., for example, assumed constant excitation through pictures to model transitions between emotions [162].

While this assumption is not realistic, it is generally not clear when and how strong (i.e. the development over time) such a single picture actually induces an emotion. The study by Bailenson et al. avoids the issue of subjective feeling self-report by annotating the data by a professional based on facial expressions [172].

When the desired labels should represent the subjective feeling, these common approaches fail to deliver. One option used by Wen et al. is to use a button (with two states: on/off) and to ask the user to mark the periods of strong experienced emotional episodes in a re-run of the experiment [160]. This is based on the assumption that the subject can remember their emotional experience and that experienced emotions were "strong feelings", limiting the labels to be dichotomous.

To measure the affective experience continuously in dimensional space (e.g. VA space) and time, several possibilities are mentioned in literature. Dials and sliders are often used in the field of market research, and found application to the assessment of affective states as well. Laurans et al. developed a physical device called *emotion slider*, which can collect self-reports of users in one dimension by pushing or pulling the knob of the slider in the appropriate direction [177]. Software solutions that use a computer mouse or a joystick as input devices are also frequently used. An early development is *Feeltrace* which was developed by Cowie et al. in 2000 [178]. It displays the position of the currently selected affective state in evaluation-activation space together with changing color depending on the position and decreasing size over time, i.e. leaving a trace behind. *Gtrace* is a more recent and flexible version of it [179]. *EMuJoy*, developed by Nagel et al., is functionally very similar but additionally allows for synchronization and control of both auditory and visual stimuli [180].

Still, there are some limitations to the currently used dynamic measures. First of all, online annotation poses a more demanding task than post-assessment that may require some practice. Tracing two dimensions simultaneously will have a negative effect on the quality of tracing; more than two dimensions yield no satisfactory results [19]. At the same time, using only one dimension of the VA space significantly limits the scope of experiments. Further, annotator agreement in continuous time labels needs special attention. While moments of the change in annotations (relative values) is quite stable across subjects, the absolute value varies a lot due to a person's internal scale. This has been shown in a study where subjects were to continuously label actor recordings [181].

5.3 Approach

While emotion dynamics have been addressed on various levels, current approaches to emotion recognition from physiological signals do not address the fact that emotions are dynamic and, thus, require dynamic models for prediction of the affective state.

The CPM itself is very descriptive as in which components exist and interact within the emotion process. Figure 5.1 shows the mechanism of component integration from Scherer's CPM, giving the subjective feeling a special role in this setup [32]. Changes in appraisal lead to changes in other components, e.g. ANS. These changes influence our subjective feeling which, in turn, regulates our appraisal. The CPM makes detailed predictions on how appraisal processes, i.e. outcomes in SECs as listed in the first two columns of Table 5.1, effect other components.
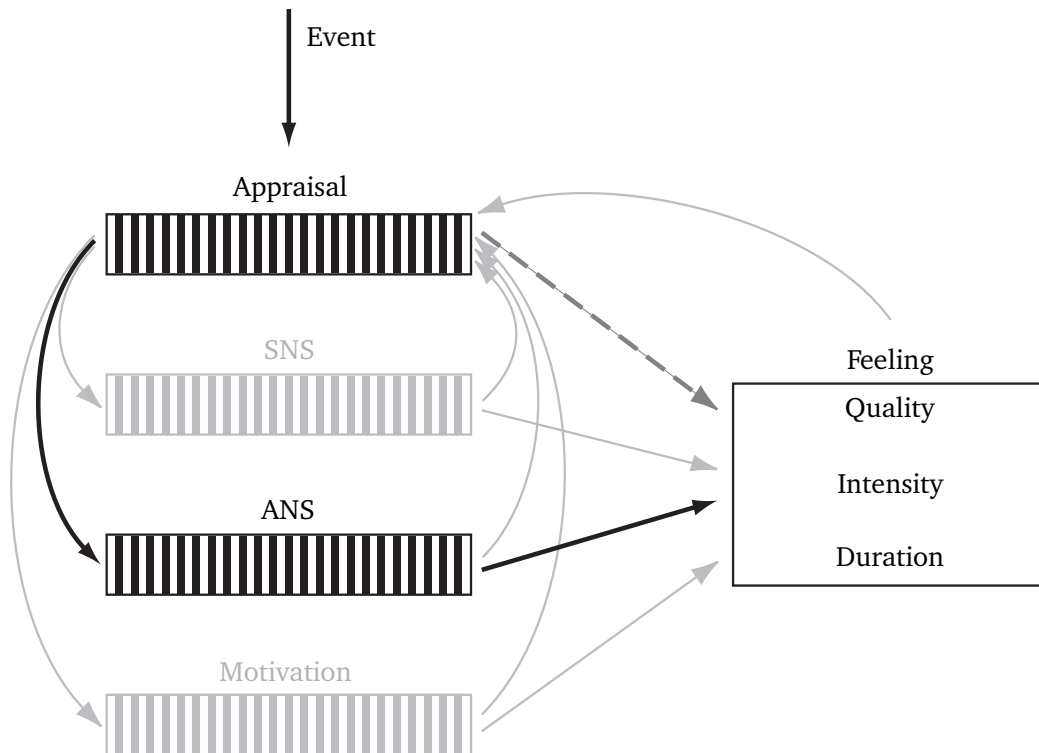
Figure 5.1: Mechanism of component integration (redrawn based on [32])

Looking at the model from an emotion recognition point of view, we want to infer the current actual state of subjective feeling, which is characterized by the intensity I of emotion quality ϑ at time t and can generally be written as $I(\vartheta, t)$. An event that may cause a change in the feeling component will set in motion and affect the outcome of appraisal processes. These, in turn, influence and interact with all other components including the feeling component. However, appraisal changes are not directly measurable and, hence, cannot be used for emotion recognition directly (dark gray dashed arrow). But we can measure changes in ANS in reaction to SECs. The challenge is to establish a mapping from changes in ANS to changes in the feeling component (black solid arrows). This has obviously been attempted before (see section 5.2.3), but has left out the consideration of dynamics. What is missing is a dynamic model that combines the parts of the subjective feeling state and its history.

Similar to Bailenson et al. [172], we divide the subjective feeling component into separate estimation of emotion quality $\hat{\vartheta}(t)$ and emotion intensity $\hat{I}(\vartheta, t)$ as shown in Fig. 5.2. But here, the quality estimate $\hat{\vartheta}(t)$ is provided as an input to the Emotion Intensity block, since the complete affective state is characterized by quality *and* intensity over time.

One repeatedly reported issue is the many-to-one mapping problem when trying to predict an affective state from physiological signals. This points to the fact that unknown other factors could also influence the physiological response signal [182]. Thus, in order to understand emotion processes, they have to be carefully controlled for in the experiment design, such that we can assume no other external stimuli are present that influence the signal of interest. But even if all ANS changes are related to the emotion process, they might still correspond to different SECs, which interact non-linearly with other components, as the CPM proposes. As an example, Table 5.1 lists the effect of SECs on a physiological signal, here
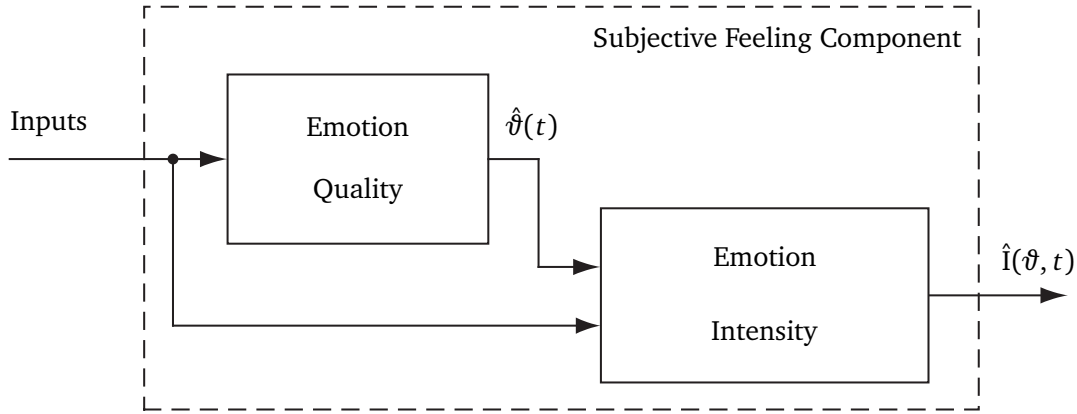
Figure 5.2: The subjective feeling component can be divided into separate, consecutive estimation of emotion quality $\hat{\vartheta}(t)$ and emotion intensity $\hat{I}(\vartheta, t)$.

GSR. Many evaluation checks are predicted to result in an increase of GSR, but probably their influence on the feeling component varies strongly. This illustrates the need for a non-linear model that can take time evolution into account.

By utilizing expected changes in SECs together with predictions of their effect on ANS to carefully design experiments, a systematic study is possible. And by additionally collecting subjective feeling data together with physiological reactions, we can model their dynamic influence on the subjective feeling component. Herein, we believe that combining both the understanding gained through the appraisal processes with data-driven methods for emotion recognition can greatly enhance the quality of estimation, as it has also been advocated by Mortillaro et al. [34]. While white-box approaches require exact theoretical analysis as well as a known model structure and correspondence of parameters to physical variables, purely data-driven black-box approaches from non-linear system identification merely attempt an input/output mapping but lack interpretability [183]. In contrast, gray-box approaches can incorporate basic knowledge from theory, e.g. structure and interactions, and allow meaningful interpretation of parameters while at the same time allowing for flexibility of experimental analysis. In validating the internal consistency of the model or finding the mapping from component dimensions to discrete emotion labels, this approach combining theory with experimental data has been successfully applied before (e.g. [161], [184]).

The Dynamic Field Theory (DFT) introduced in the next section provides a biologically-inspired gray-box approach suitable for emotion recognition. In this thesis, we limit ourselves to the modeling of the Emotion Intensity block and assume the emotion quality estimates $\hat{\vartheta}(t)$ to be defined by experimental design. An extension of the proposed dynamic model for emotion quality estimation is straightforward, but its benefits over existing methods, e.g. common machine learning methods, for this case would need to be proven elsewhere. We propose a dynamic architecture for the estimation of the subjective feeling state that can be trained with user-data and used to make predictions from multiple modalities. The main contribution lies in the capability to model non-linear effects of changes in ANS to estimate emotion intensity. This includes merging the current emotion quality estimates with the history of the complete subjective feeling state $\hat{I}(\vartheta, t)$.

Tabular 5.1: Effects of appraisal evaluation checks on skin conductance (from Scherer, 2009 [32]).

stimulus evaluation check (SECs)	outcome	skin conductance response (GSR)
relevance	novel and goal relevant	increase
	pleasant	-
	unpleasant	increase
implications	conductive	-
	obstructive	increase
coping potential	no or low control	-
	high control/high power	-
	control possible/ low power	increase (secretion of sweat)
normative significance	requirements met or surpassed	ergotropic shift plus elements of pleasantness and high power response
	incompatible	ergotropic shift plus elements of unpleasantness and low power response

While it is possible to control for the kind of emotion quality induced during an experiment, every subject may have a subjectively different experience of intensity both in temporal evolution and internal scale. The decomposition of emotion quality ϑ and intensity I in accordance with the emotion dynamics of type 3 and 4 (as introduced in Section 5.2.2) has advantages on the practical side, too. It facilitates the systematic study of dynamics by allowing us to focus on studying intensity dynamics independently from quality dynamics, i.e. we assume the latter to be defined by experimental design. Thereby, we can dynamically measure the affective state in only one dimension, i.e. emotion intensity, while at the same time not posing limits on the experimental variety of emotion qualities. Further, tracing intensity of an emotion has been shown to have a high reliability between subjects [19].

We further adopt the representation of affective state as quality and intensity and its relation to the VA space using polar coordinates as a way to relate emotion qualities to each other. While our approach is not limited to this particular angle-based distance-measure, we see great advantages in being able to model emotion quality relations on a continuous scale. At the same time, this relation allows the usage of common, well-established induction methods such as the International Affective Picture System (IAPS) when conducting experiments, such as the one presented in Chapter 6.

5.4 Dynamic model

We briefly introduce the main idea and basic concepts of Dynamic Field Theory (DFT), which is a well established model for embodied human cognition [185]. The dynamic model for emotion intensity estimation is inspired by this framework and is introduced subsequently.

5.4.1 Dynamic Field Theory

DFT can be seen as a generalization of recurrent neural networks to continuous dimensions, adding a functional interpretation to each layer. Such dimensions typically represent behavioral states, such as location or perceptual features, and are called Dynamic Neural Fields (DNFs) [186]. A DNF evolves over time governed by the following equations, which have first been analyzed by Amari in 1977 [187]:

$$\tau_u \dot{u}(x,t) = -u(x,t) + h_u + S_u(x,t)$$
$$+ \int f_u(u(x',t))\omega_u(x-x')dx' ,\tag{5.1}$$

$$\omega_u(x-x') = c_{exc} \exp\left[-\frac{(x-x')^2}{2\sigma_{exc}^2}\right]$$
$$-c_{inh} \exp\left[-\frac{(x-x')^2}{2\sigma_{inh}^2}\right] ,\tag{5.2}$$

$$f_u(u(x,t)) = \frac{1}{1 + \exp\left[-\beta_u u(x,t)\right]} .\tag{5.3}$$

In (5.1), $u(x,t)$ is the activation of the DNF over dimension x and time t. The second term h_u denotes a negative resting level; $S_u(x,t)$ is the input function. The integral term defines the lateral interactions (e.g. local excitations and global inhibition within the DNF), which includes the interaction kernel $\omega_u(x-x')$, (5.2), and the sigmoidal non-linearity $f_u(u(x,t))$, (5.3), which is the output of the DNF. Parameters c_{inh}, c_{exc}, σ_{inh}, σ_{exc} are the amplitude and width of the inhibitory and excitatory part of the interaction kernel, respectively, and β_u controls the slope of the sigmoid. The rate of relaxation of a DNF is given by the time constant τ_u.

In multilayer designs, couplings between layers is possible by, for example, using weighted, additive inputs to a DNF as shown in (5.4). For a field v coupled with u from (5.1) this results in:

$$\tau_v \dot{v}(y,t) = -v(y,t) + h_v + \int f(v(y',t))\omega_v(y-y')dx'$$
$$+ W(x,y)f_u(u(x,t)) ,\tag{5.4}$$

where $W(x,y)$ represents entries of a weighting matrix, which maps the dimensions of space x onto the dimensions of the space y [186]. Other ways of connecting and combining layers as well as layers with special functionality have been developed but are not required here for now. For more information on the DFT and its mechanisms as well as relations of the DFT

to other neuromorphic architectures, the reader is kindly referred to a recent and detailed review by Sandamirskaya [186].

5.4.2 DFT for intensity estimation

Using the concept of DFT introduced above, we propose a multi-layer DNF for intensity estimation, here introduced in its simplest version using three layers as shown in Fig. 5.3a. All fields are defined over the space of emotion quality ϑ. The internal field dynamics are defined by the interaction kernel $\omega(\vartheta - \vartheta')$ of each field.

An input layer u (according to (5.1)) represents affective cues given by changes in ANS signals predicted from SECs, e.g. the change in GSR. The function of such a layer is to recognize events that influence the affective state and direct this information by merging it with the emotion quality estimate $\hat{\vartheta}(t)$. To simplify, we only use one input layer here. But generally, these can (and should) be extended to other inputs as well. This layer has only unidirectional (forward) interaction with the following two layers i and m, but different input layers can interact with each other bidirectionally to model interactions. The input to such a layer is the quality estimate $\hat{\vartheta}(t)$ combined with the activity of the layer's predicted SEC effects. The quality estimate $\hat{\vartheta}(t)$ can take different forms (e.g. probabilities over the whole range of ϑ or one or more discrete values) depending on the method implemented and output given by the Emotion Quality block. If the output is one discrete value, which we assume here without loss of generality, integration into the field can be achieved, for example, by:

$$S_u(\vartheta, t) = g(\vartheta \,|\, \hat{\vartheta}(t), \sigma_u) + \tilde{A}(t) \,, \tag{5.5}$$

where $g(\vartheta \,|\, \hat{\vartheta}(t), \sigma_u)$ denotes a Gaussian function over the field of ϑ with estimated emotion quality $\hat{\vartheta}(t)$ as mean and an uncertainty measure σ_u as standard deviation (similar to Gaussian Processes [188]). The feature describing the predicted SEC effect (e.g. GSR change), denoted $\tilde{A}(t)$, is used as a boost of resting level to this layer.

The core part of the model is the intensity layer i, which follows the equation:

$$\tau_i \dot{i}(\vartheta, t) = -i(\vartheta, t) + h_i + S_i(\vartheta, t)$$
$$+ \int f_i(i(\vartheta, t)) \omega_i(\vartheta - \vartheta') d\vartheta' \,. \tag{5.6}$$

It is coupled with both u and m as shown in (5.4), summarized as the additive inputs $S_i(\vartheta, t)$. Activity at each location of the DNF, i.e. emotion quality ϑ, is related to the emotion intensity. The output of i is defined as the emotion intensity estimate $\hat{I}$, given as:

$$\hat{I}(\vartheta, t) = f_i(i(\vartheta, t)) \,. \tag{5.7}$$

Additionally, a working memory layer m is added and coupled with u and i, following the same structure as given in (5.6) for layer i. Its main function is to stabilize intensity-peaks in field i.
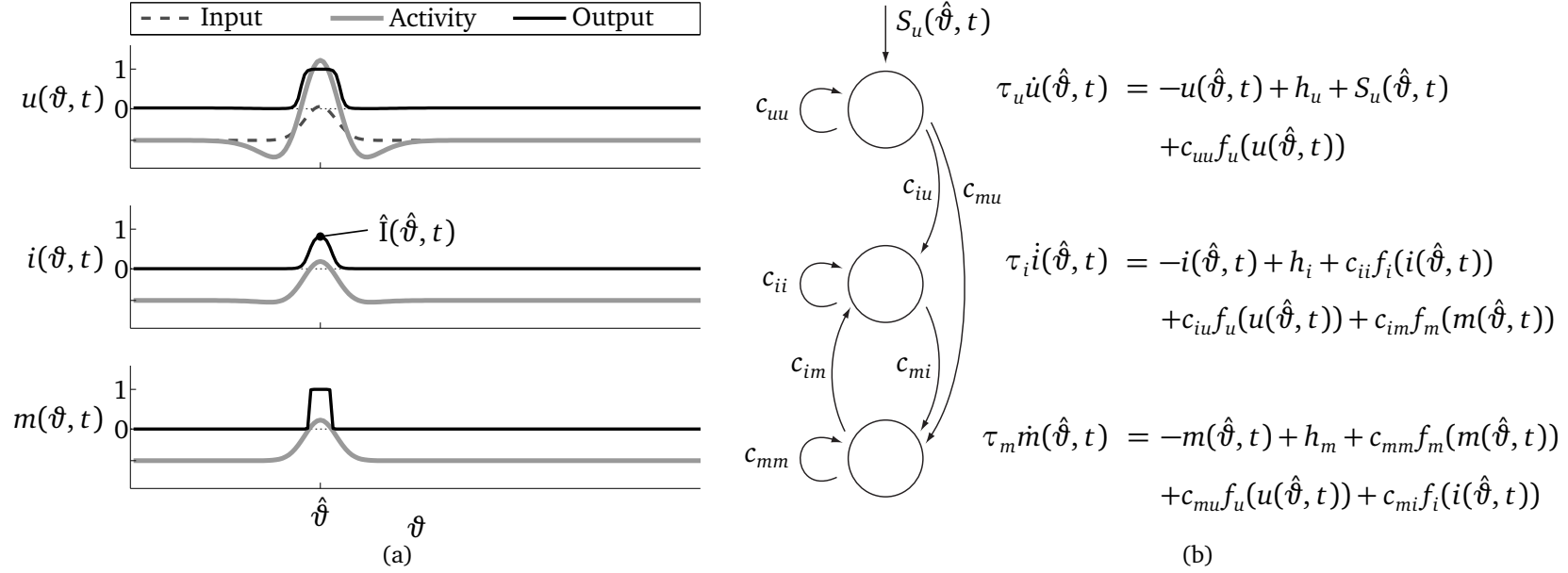
Figure 5.3: Architecture of the proposed dynamic model: (a) three-layer field over the dimension of emotion quality ϑ and (b) the equations at a single location $\hat{\vartheta}$ of the field.

The architecture above for integration of appraisal dimensions into intensity estimation constitutes the theory-based part of the approach. It presents a flexible framework that allows systematic analysis of the mapping between changes in ANS predicted by SECs and the subjective feeling component. The model is trained using physiological data and subjective feeling data from subjects to optimize estimation (see Sec. 6.1.2). The layered design further allows to incorporate multi-modal signals from other components or context information.

In the following we discuss some properties of the model and introduce a concrete application of this framework to estimate emotion intensity for a single emotion quality.

5.4.3 Properties of the model

Different forms of emotion dynamics introduced in Section 5.2.2 (type 1-4) are considered in the proposed model. Herein, knowledge about type 1 dynamics can be included in the architectural structure of the layers and interactions. Although type 2 dynamics such as mood states are not part of the Emotion Intensity block, they can easily be factored in, for example, by adding a long-term memory layer to the model. Some forms of type 3 dynamics, such as the trajectories of $\hat{\mathrm{I}}(\vartheta, t)$ between two consecutive emotion qualities, depend on the history of the whole subjective feeling state. These *lateral dynamics* can be influenced by means of the parameters of the interaction kernel. This way, transition trajectories as investigated by Hakim et al. [163] seem to intuitively be reproduced by this model by exploiting the design of the interaction kernel. A formal proof is yet necessary, but remains future work. Other forms of type 3 dynamics, such as the transition probabilities from the current emotion quality to the next, have to be considered in the Emotion Quality block. Type 4 dynamics refer to changes in the output of layer i and are of special interest in this part of the thesis. We take a closer look at this topic in the next section.

An advantage of using the angle as distance measure is that the field is circular, therefore allowing to relate all emotion qualities to each other when computing the folding-integral with the interaction kernel. Although there is quite some evidence in favor of considering a circumplex model for the quality of emotion (e.g. [27], [30]), the continuous property of ϑ is not a necessary condition. The model also covers discrete emotion representations by using ordinal nodes for discrete emotion qualities, preserving the benefits of the layers functional meaning.

5.4.4 Intensity estimation for a single emotion quality

The decoupling of subjective feeling makes it easy to isolate the question of intensity changes at a fixed emotion quality $\hat{\vartheta}$, i.e. type 4 dynamics. This allows studying a well-defined part through proper experimental design and later transferring the single findings to the whole field. In this case, the model simplifies to the equations given in Fig. 5.3b. The input to field u at a single location $\hat{\vartheta}$, $S_u(\hat{\vartheta}, t)$, is a scalar of only the predicted SEC activity $\tilde{\mathrm{A}}(t)$; the constant contribution of $g(\hat{\vartheta}\,|\,\hat{\vartheta}, \sigma_u)$ is included in the resting level h_u. The parameters that need to be chosen for this model are the time constants for each layer, τ_u, τ_i, and τ_m, resting levels h_u, h_i, and h_m, and the slope parameters of the output functions, β_u, β_i and β_m. Coupling and interaction are defined by self-excitation/inhibition weights c_{uu}, c_{mm} and c_{ii} as well as

coupled additive input weights c_{iu}, c_{im}, c_{mi} and c_{mu}. The data-driven method to identify these is explained in Section 6.1.2.

The following *zero-class* property is noteworthy: if no emotion quality is estimated (or only with very high uncertainty), then the input layer is not excited enough by the boost of $\tilde{A}(t)$ and no intensity changes are predicted, i.e. there is no non-zero output of any affective state.

5.5 Conclusion

Emotion theories such as the CPM have pointed out the dynamic nature of emotion processes making it necessary to revise the approaches used for emotion recognition, too. In this chapter, we have established an approach to incorporate appraisal into emotion recognition by combining existing emotion theory with data-driven methods.

For the first time, a dynamic gray-box model framework based on DFT has been proposed that is capable of finding mappings from physiological signals to affective states while taking predictions from stimulus evaluations checks into account. Parameters of the model can be identified using experimental data while a functional meaning of each layer is sustained. We discussed the main properties of the model as well as an in-detail version of the model for estimating emotion intensity for a single emotion quality. Importantly, it provides the ability to study a well-defined part of the field while retaining a way to combine results in one model. This case is further investigated in the next chapter by means of an exemplary study to estimate emotion intensity from GSR for a fixed emotion quality.

Future work includes the design of the interaction kernel when considering lateral dynamics as well as several conceivable extensions of the model further elaborated in Section 6.5. We limited ourself to model the Emotion Intensity block here assuming the emotion quality to be fixed by experimental design. Further development on the modeling of Emotion Quality estimation and its interaction with Emotion Intensity would be interesting.

6 Emotion intensity estimation from physiological signals

Summary. *The dynamic model proposed above facilitates controlled experimental designs to study emotion dynamics. A proof-of-concept study is presented in this chapter showing how experiments can be conducted and evaluated. We find that the dynamic model performs significantly better than baselines. In summary, the main contributions of this chapter are*

- *the design and performance of an exemplary study measuring emotion intensity levels over time in combination with physiological signals,*

- *a comparison of evaluation measures in continuous emotion recognition,*

- *an implementation and evaluation of the dynamic model to estimate emotion intensity from changes in physiological signals.*

The dynamic framework introduced in the previous chapter establishes a possibility to integrate knowledge about the emotion elicitation process into an emotion recognition system. By allowing us to focus on a well-defined part of the dynamic evolvement of the affective state, it provides the means to shed light onto this hitherto neglected part of the emotion recognition.

The main contributions of this chapter are concerned with carrying out a proof-of-concept study testing the in-detail version of the model given in Fig. 5.3b and discussed in Sec. 5.4.4. In this, we show how appropriate experiments can be designed and we investigate the model's capabilities and limitations on a real-world dataset. Therefore, a dataset to dynamically estimate intensity of negative emotions from physiological signals is recorded. We test the proposed model on this dataset and compare it to a common static approach to emotion recognition as well as two baselines. In evaluating intensity estimates on a continuous scale instead of discrete classes, different evaluation measures are applied. To make an informed choice, we review and discuss the suitability of various measures that have been proposed in the context of emotion recognition in dimensional space.

The remainder of this chapter is organized as follows: Section 6.1 introduces setup and design of an exemplary user study to estimate emotion intensity from physiological signals using the proposed dynamic model. A detailed discussion of evaluation measures for emotion recognition using continuous scales is presented in Section 6.2. Results of the user study and a discussion are given in sections 6.3 and 6.4, respectively. We conclude with a summary, limitations and outlook of this model in Section 6.5.

6.1 Experiment and data acquisition

As an example of the mapping from physiological signals to subjective feeling using the proposed model, we study the possibility to predict emotion intensity from changes in GSR over time for a single emotion quality. We chose GSR, since it is an easily accessible signal that is predicted to have the same reaction, i.e. *increase,* to multiple SECs (see Table 5.1). We describe the design and details of a study carried out to collect data in this section.

6.1.1 Setup and data processing

The experiment used for data acquisition was set up in a closed room controlled for sound and light conditions. The general setup used for data acquisition is visualized in a schematic way in Fig. 6.1. Induction was based on display of emotion pictures on a screen (screen 1) in front of the subject, who was seated in a comfortable chair with arm rests. As input devices, the subject used a mouse and a slider device. The latter was developed in-house based on a linear potentiometer which was connected to the recording PC via an Arduino Uno USB interface (see also Appendix A.1.3) and was positioned on the right arm rest. It was used to collect online emotion intensity ratings at 8 Hz and was integrated into the physio data stream using a zero-order hold. The slider position was normalized to range from $[0; 1]$ and corresponds to the subjectively felt intensity $I(t)$. During the experiment, the subject could monitor the current slider position on a second screen (screen 2). Additionally, synchronization messages via UDP between induction and recording PC were sent to facilitate proper segmentation of trials.

A sensor for GSR, attached to middle and ring fingers of the left hand of the subject, was connected to the recording PC via g.tec gUSBamp. The amplifier was set to include a 30 Hz low-pass filter and recorded at 256 Hz. The first difference of the GSR signal,

$$\delta_{GSR}(t) = \xi_{GSR}(t) - \xi_{GSR}(t-1) \,, \tag{6.1}$$

was extracted as input to the neuron/layer u of the dynamic model, down-sampled to 25.6 Hz. Since we assumed a fixed emotion quality $\hat{\vartheta}$, this resulted in an input $S_u(\hat{\vartheta}, t) = \delta_{GSR}(t)$.

6.1.2 Parameter identification

All parameters, listed in Section 5.4.4, are combined into a vector **p**. We define the function $\hat{I}(t) = \text{simDynModel}(S_u(\hat{\vartheta}, t), \mathbf{p})$ that simulates the dynamic model for a given input time series $S_u(\hat{\vartheta}, t)$, $t = 1, \ldots, T$ and returns the emotion intensity estimate $\hat{I}(t)$ for each point in time. Differential equations were implemented by means of difference equations. All initial conditions are set to zero.

The free parameters **p** of the model can be identified such that they minimize the mean squared error (MSE) between the estimated intensity $\hat{I}(t)$ and the subjective intensity labels
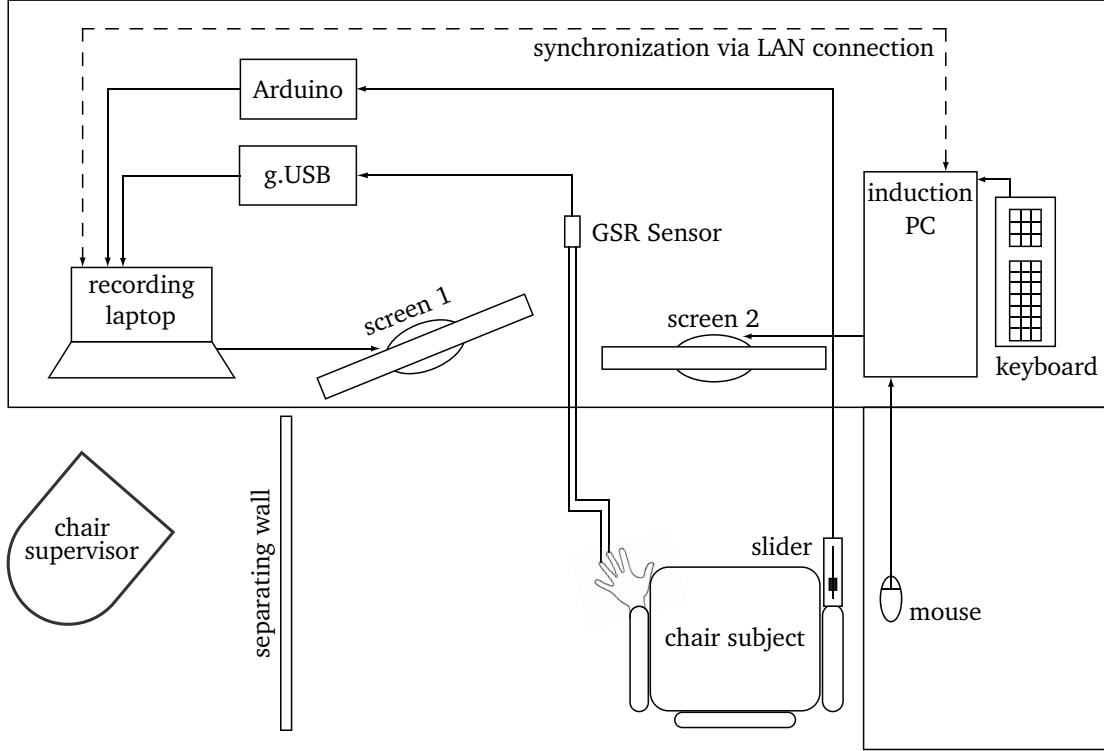
Figure 6.1: Experiment setup

I(t) recorded during the experiment. Therefore, we define the optimization problem:

$$\min_{\mathbf{p}} \frac{1}{T} \sum_t \left| \hat{\mathrm{I}}(t) - \mathrm{I}(t) \right|^2 \tag{6.2}$$

$$\text{s.t. simDynModel}(\mathbf{0}, \mathbf{p}) < \epsilon \,,$$

$$\mathbf{p}_{min} < \mathbf{p} < \mathbf{p}_{max} \,,$$

where $\mathbf{p}_{min/max}$ define lower and upper bounds of the parameters (see Appendix A.2), respectively, and the inequality constraint assures that the model output is smaller than ϵ, given zero input, denoted as $\mathbf{0}$. Here, we set $\epsilon = .05$.

Since non-linear models with such a high number of free parameters are prone to overfitting, leave-one-out cross validation was chosen to train and evaluate the proposed model[1]. When parameters are identified by means of multiple trials $n = 1, \ldots, N$, the median of all MSE_n is used as optimization criterion. This results in a more balanced model and puts less emphasis on outliers than using the average over MSE_n as a criterion. The parameters estimated for the presented results are provided in the Appendix A.2.

---

[1]In system identification, the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) are often used for model assessment. Cross-validation, which is well-established in machine learning applications, is probably the simplest and most widely used alternative to AIC and BIC [52, Chap. 7]. The interested reader is pointed to Stone, who showed that the leave-one-out cross-validation is asymptotically equivalent to AIC [189], and to further discussion by Watanabe [190].
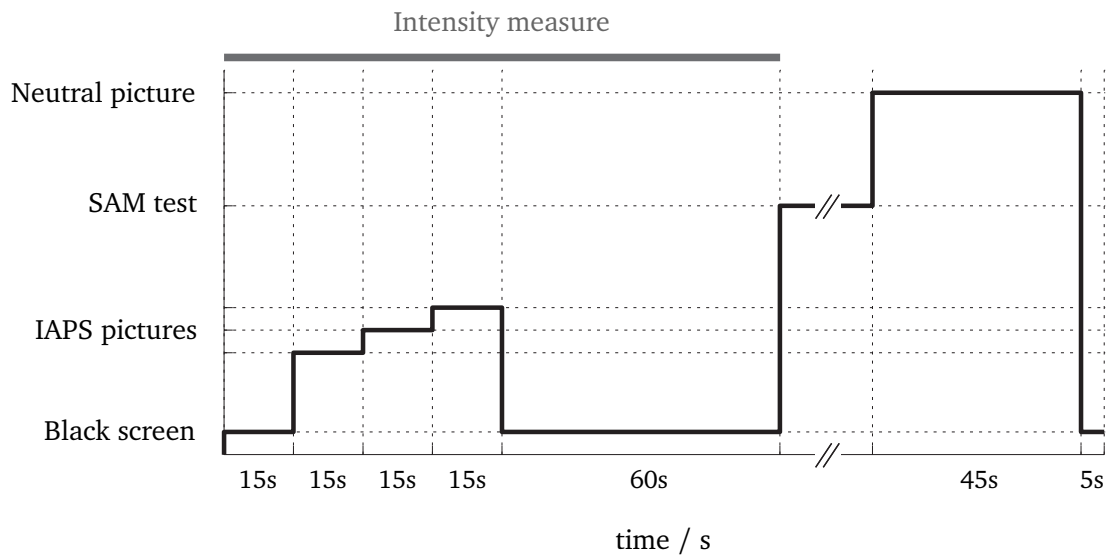
Figure 6.2: The emotion induction protocol for each trial consisting of the set of IAPS pictures, recording time before and after the pictures, SAM test, and a neutral picture. Additionally, the time in which the intensity measure was taken is indicated.

6.1.3 Protocol

Emotions were induced using IAPS pictures [41]. We selected pictures on the basis of dimensional ratings, that are included in the IAPS database, by means of an in-house developed tool to find best matches within a user-defined range with minimal variance. Three high-intensity regions of emotion qualities were chosen for induction in order to balance the experiment: the affective states *positive, negative* (both with added arousal), and *calm*. Due to the known variance in response to visual affective stimuli between male and female subjects [41], we selected two separate picture sets to induce emotions more reliably.

Figure 6.2 shows the induction protocol implemented. Each trial started with a black screen with fixation cross, followed by 3 randomly drawn pictures from a single emotion quality, each shown consecutively for 15 seconds. Every picture was used only once per subject. To capture the decay of emotion intensity after the stimulus was gone, recording of each trial was prolonged by 60 s showing a black screen with fixation cross. During this entire time, the subjectively experienced emotion intensity was recorded. Afterwards, a SAM questionnaire (Self-Assessment-Manikin Test [45]) was filled out for the dimensions valence and arousal on a scale from [1; 9] with the purpose of post-evaluation of induction reported in Section 6.1.5. Each trial ended with a neutral picture (shown for 45 s) to help the subject reach a neutral state again and to reduce effects of transition probability [157]. In total, 21 trials were recorded for each subject (7 for each emotion quality), while the order of emotions was randomized. To avoid effects of beginning and ending of trial, five seconds were cut off each end for further analysis.
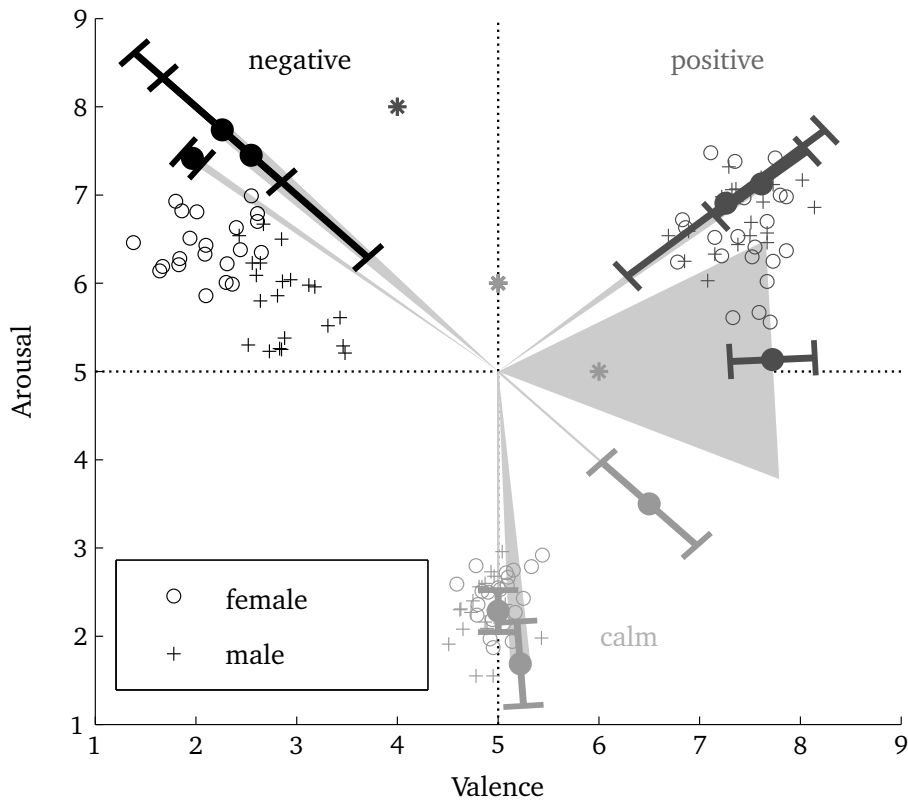
Figure 6.3: Valence-Arousal values of IAPS picture sets (targeted states) together with mean and variance of the SAM tests for each subject (excluding outliers marked as asterisks).

6.1.4 Subjects

Three subjects aging between 29 and 35 participated in the experiment. Two of them have engaged in emotion studies once before. Ethical approval for this experiment was obtained from the ethics committee of the medical faculty of TUM. Subjects were instructed on the experiment procedure and filled out a questionnaire testing for suitability (e.g. skin allergies), well-being (WHO), and trauma screening (PTSD). No subject had to be excluded.

Subjects were instructed to focus on the subjectively felt intensity of emotion and indicate this using the slider device, where intensity equals deviation from neutral state. The picture should help to guide the development of the quality of emotion. They were asked to imagine themselves at the place and time of the moment the picture was taken. Further, the SAM test was explained and they were asked to indicate the overall experience of a trial when assessing their affective state.

After being equipped with recording sensors, a test-run (using an additional set of pictures) was conducted prior to the actual experiment to accustom the subjects to the protocol and the task of directing attention to the inside while simultaneously adjusting the slider position. Especially the latter required some practice for subjects.

6.1.5 Evaluation of induction

We used the SAM test results to check whether the intended emotion was successfully induced. Figure 6.3 shows the targeted areas of induced emotions in the valence-arousal space given by IAPS annotations of the picture sets (circles and plus signs). In addition, the mean (vector) and variance (bars) of the SAM tests were transformed to polar coordinates, i.e. angle ϑ and length of vector I, and plotted for each induced emotion quality and each subject. We selected *negative* emotion quality for further analysis, since induction of emotion quality was most consistent here, i.e. we see the smallest variance of angle ϑ. But the exemplary study could be carried out for the other emotion qualities as well. With one exception, the variance of subjectively felt emotion quality was generally lower than the range of experienced emotion intensity, which we interpret as evidence in favor of our experiment design.

6.2 Evaluation measures

When predicting emotional states on a continuous scale, e.g. the intensity or in VAD space, classically used quality measures from classification problems like accuracy are not easily applicable. Thus, a series of alternative measures have been proposed in emotion classification literature. The most important ones are introduced below.

1) In an effort to directly transfer accuracy from discrete classes to the evaluation of dimensional approaches, the scale is often divided into discrete areas. For example, Yang et al. measure accuracy $\text{A}_{discrete}$ of their regression approach by comparing the sign of the prediction $\hat{y}_i$ with the sign of the ground truth y_i, which in this case ranged from $[-1; 1]$ [191]:

$$\text{A}_{discrete} = \frac{1}{N_T} \sum_i I(\hat{y}_i, y_i) \tag{6.3}$$

where

$$I(\hat{y}_i, y_i) = \begin{cases} 1 & \text{if } \hat{y}_i y_i > 0 \\ 0 & \text{else,} \end{cases} \tag{6.4}$$

and N_T is the total number of samples. This results, however, in a rather coarse evaluation, if the reason for using the dimensional model is to detect and process more subtle changes in emotion.

2) A step further, Haag et al. proposed a modified accuracy measure which counts those samples as *correct* that are within a certain bandwidth b, e.g. 10%, of the target value.

$$\text{A}_{bdw} = \frac{1}{N_T} \sum_i I(\hat{y}_i, y_i) \tag{6.5}$$

where

$$I(\hat{y}_i, y_i) = \begin{cases} 1 & \text{if } \|\hat{y}_i - y_i\| < \text{b}(y_{\max} - y_{\min}) \\ 0 & \text{else,} \end{cases} \tag{6.6}$$

where $\|\cdot\|$ denotes the Euclidean distance. The threshold of distance includes a normalization for scale, computed from its maximum and minimum value y_{max} and y_{min}, respectively. Which bandwidth is generally acceptable has to be determined for each given application scenario individually.

3) Most studies report the mean squared error (MSE) or mean absolute error (MAE) between the estimated emotion $\hat{y}_i$ and its ground truth y_i [141], [175], [192]:

$$\text{MSE} = \frac{1}{N_T} \sum_i (\hat{y}_i - y_i)^2 \, , \tag{6.7}$$

$$\text{MAE} = \frac{1}{N_T} \sum_i |\hat{y}_i - y_i| \, . \tag{6.8}$$

Note, that these measures are not normalized. Thus, sometimes the MSE/MAE of a random regressor is given as a baseline for a more meaningful interpretation. Yet, it is often not intuitive to compare different studies, since the numerical range of the adopted dimensions varies.

4) Another possibility is to report R^2 statistics (as done in [191]), which represents the portion of the target variable variation explained by the regressor. There are multiple definitions for R^2 depending on the regression type; for linear regression R^2 is defined as:

$$R^2 = \frac{SS_{reg}}{SS_{tot}} = \frac{\sum_i (o_i - \mu)^2}{\sum_i (y_i - \mu)^2} \, , \tag{6.9}$$

where SS_{reg} denotes the regression sum of squares and SS_{tot} the total sum of squares. The associated model value for each training value y_i is denoted by o_i; μ is the mean of all training values. The R^2 statistic is calculated to asses the quality of a linear regression model with respect to the training data and ranges from $[0; 1]$, independent of the range of target values. In machine learning problems, however, we are interested in the ability of a model to generalize, and thus, in the performance given on a test data set.

5) Last, the correlation coefficient r between the estimates and ground truth has been used in some studies, e.g. [172], [173]. It is defined as:

$$r = \frac{1}{N_T \sigma \hat{\sigma}} \sum_{i=1}^{N_T} (y_i - \mu)(\hat{y}_i - \hat{\mu}) \, , \tag{6.10}$$

where σ and $\hat{\sigma}$ are the variances and μ and $\hat{\mu}$ are the means of all $\mathbf{y}_i$ and $\hat{y}_i$, respectively. But r only measures the degree of linear relationship between y and $\hat{y}$ and, thus, does not account for the scale of the prediction, i.e. bad absolute estimations can still yield high values for r, which makes this measure unsuitable for the proposed purpose in our opinion.

Summarizing, measure 1), the discretized accuracy $A_{discrete}$, only gives a coarse evaluation, measure 4), R^2, is more suitable to evaluate the quality of the regression model, rather than to evaluate prediction ability, and measure 5), r, can be ambiguous regarding the absolute error of the prediction. Hence, for emotion recognition in continuous dimensional space, mean squared error (MSE) and bandwidth accuracy A_{bdw} have been selected as objec-

tive evaluation measures. We see these as the two options most used and with high potential to provide a more general evaluation measure that allows a systematic comparison between studies (for more details, see also [203]).

We computed A_{bdw} for b $= \{.1, .15, .2\}$, i.e. estimations within $\pm 10/15/20\%$ of the actual intensity are considered *correct* classifications. Additionally, we compare our results to a linear ridge regression and two baseline random estimators. Linear ridge regression is fitted from both $\xi_{GSR}(t)$ and $\delta_{GSR}(t)$ at each point in time t (similar to the approach in [172]). The first random estimator, denoted rand-regressor, draws samples from a uniform distribution within the range of the labels, i.e. $[0; 1]$. Since the distribution of labels is not uniform, we secondly select the Π-regressor suggested by Soleymani et al. [141], which takes into account the distribution of the training labels when randomly generating intensity estimates.

6.3 Results

The results of the experiment reported in Section 6.1 are presented with the intent of answering the following questions: 1) Is there a benefit of using the proposed dynamic model over common static methods? If any, can we quantify them? 2) What are the limitations of the proposed model in its current version? How can we overcome them?

Figure 6.4 shows the intensity estimate of our model together with the subject's intensity label and the estimate using linear ridge regression for all trials and subjects. In each subplot, the model was trained using data from the other six trials. On the one hand, linear regression is clearly not suitable to estimate intensity from GSR. Since no linear correlation was found, estimates are non-zero the whole time, revealing no qualitative information. The dynamic model, on the other hand, better resembles the changes in intensity, especially at the onset of subjective feeling.

In Table 6.1 we list mean and standard deviation over all trials per subject and on average for all evaluation measures. We perform one-sided paired t-tests with Bonferroni correction comparing the dynamic model to the other implemented methods, i.e. linear regression, rand-regression, and Π-regression. The dynamic model performs significantly better than all other methods. In particular, it proofs superior to estimation using prior information about label distribution, i.e. the Π-regressor. Only three cases do not reveal a difference between the dynamic model and the compared method: Linear regression gives results similar to the dynamic model when 1) using the MSE as evaluation measure, and 2) compared to accuracy A_{bdw} at b $= .2$. And 3) no significant difference was found between the dynamic model and Π-regressor for evaluation measure A_{bdw} at b $= .1$.

A representative example showing all involved layers of the model in detail, as introduced in Section 5.4.2, is plotted in Fig. 6.5 over time. The light gray area denotes the period of induction. Activity and output of each neuron are given. In the upper plot, the changes in GSR are the input to neuron u (thin black line). Additionally, the intensity labels of the subject's rating are shown as ground truth in the middle plot.
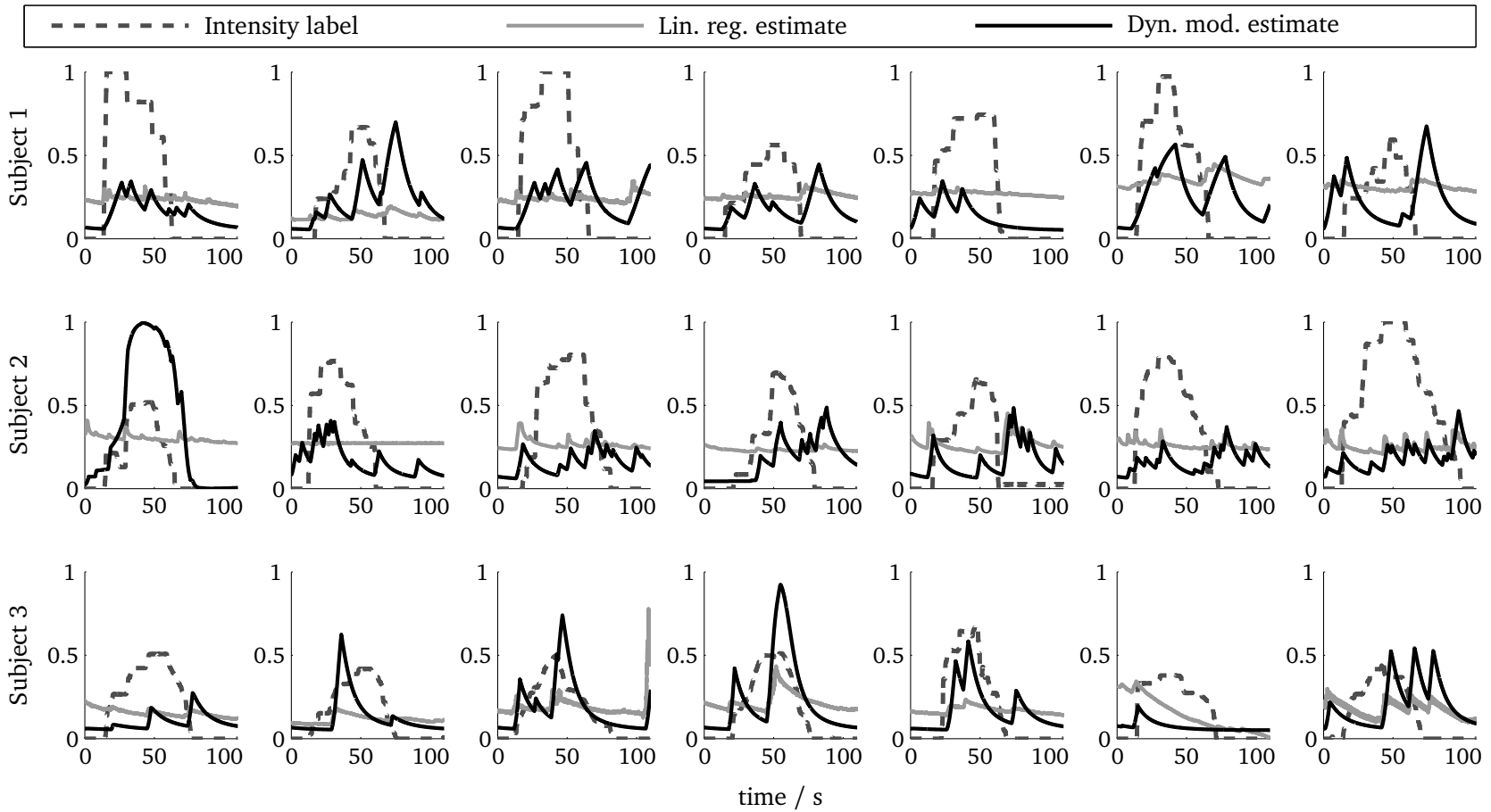
Figure 6.4: Intensity estimation for each trial and subject compared to self-reported subjective feeling (ground truth) and estimation using ridge regression.

Tabular 6.1: The mean (standard deviation) of evaluation measures MSE and A_{bdw} (b level) for the dynamic model, linear regression, and two baselines are given for each subject. Additionally, results from significance tests are given.

Model	Measure	Subject 1	Subject 2	Subject 3	Average	p-value
Dynamic model	MSE	0.11 (0.04)	0.12 (0.06)	0.03 (0.01)	0.09 (0.04)	-
	A_{bdw} (.10)	24.9% (10.8)	26.1% (12.4)	46.3% (17.4)	32.4% (13.5)	-
	A_{bdw} (.15)	37.9% (8.4)	36.9% (11.8)	60.4% (15.5)	45.1% (11.9)	-
	A_{bdw} (.20)	48.8% (7.9)	48.0% (12.8)	71.8% (11.0)	56.2% (10.6)	-
Linear regression	MSE	0.11 (0.05)	0.10 (0.05)	0.03 (0.01)	0.08 (0.04)	.79
	A_{bdw} (.10)	9.5% (8.5)	9.4% (4.1)	25.4% (15.3)	14.7% (9.3)	* .00
	A_{bdw} (.15)	17.9% (18.8)	16.7% (8.1)	45.6% (16.6)	26.8% (14.5)	* .00
	A_{bdw} (.20)	24.9% (24.3)	23.8% (8.4)	75.0% (7.7)	41.2% (13.5)	.00
rand-regressor	MSE	0.25 (0.05)	0.23 (0.05)	0.23 (0.01)	0.24 (0.01)	* .00
	A_{bdw} (.10)	14.1% (0.8)	15.2% (0.6)	15.1% (0.6)	14.8% (0.7)	* .00
	A_{bdw} (.15)	21.1% (1.2)	22.5% (0.8)	22.5% (0.8)	22.0% (0.9)	* .00
	A_{bdw} (.20)	27.9% (1.7)	29.6% (0.9)	29.6% (1.0)	29.1% (1.2)	* .00
Π-regressor	MSE	0.18 (0.04)	0.16 (0.04)	0.06 (0.01)	0.14 (0.03)	* .00
	A_{bdw} (.10)	23.9% (2.1)	22.9% (3.3)	32.0% (2.4)	26.3% (2.6)	.03
	A_{bdw} (.15)	30.3% (3.2)	30.4% (3.7)	43.7% (3.4)	34.8% (3.4)	* .00
	A_{bdw} (.20)	36.9% (4.1)	37.6% (4.3)	55.2% (3.9)	43.2% (4.1)	* .00

* denotes a significant difference to results from the dynamic model at a global level of $\alpha = .01$.
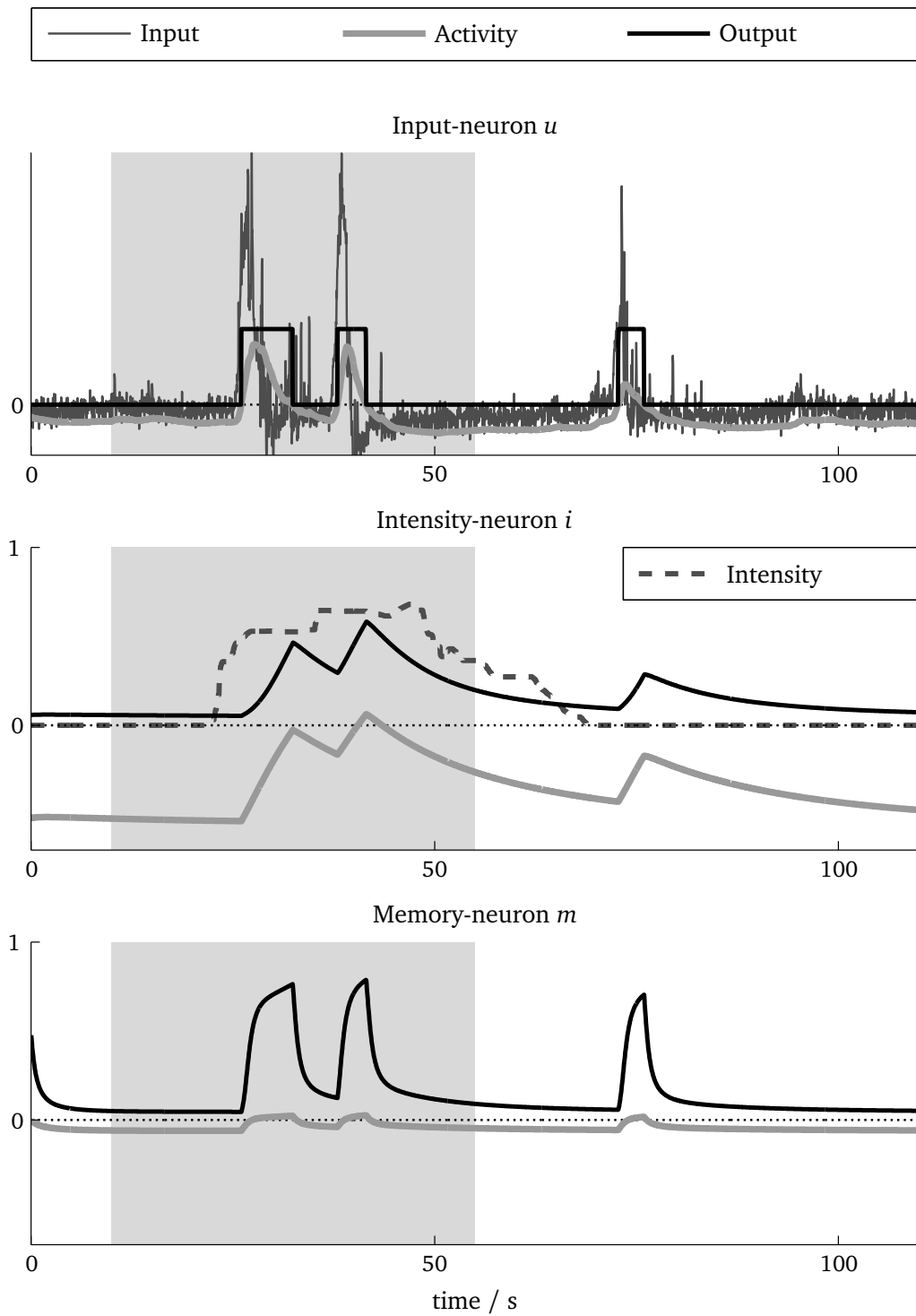
Figure 6.5: Example of all layers over time.

6.4 Discussion

Each subject responded with a relatively consistent range of intensity values to the pictures used for induction. This is in line with the finding from a study on annotator agreement by Metallinou that changes in time of the affective state are more similar between subjects than the absolute value [181]. The dynamic model is able to account for these differences in personal internal scale by tuning of parameters.

The onset of emotion intensity is often well characterized by our model (compare Fig. 6.4). Both capabilities and limitations of the proposed model are visible in the example given in Fig. 6.5. Notice that the onset and end of induction (light grey area) does not correspond to the onset and end of subjectively felt intensity (dashed line in middle plot). The input layer containing the GSR response does, however, indicate the onset as well as a second rise of intensity, which leads to a good approximation of the intensity (black solid line in middle plot). Instead of immediately decaying, the memory layer helps to stabilize the felt intensity. Although induction is over and reported intensity has decayed to zero, a third GSR response occurs around second 70. This leads to a false rise of the intensity estimation. Such GSR changes in later points in time are probably influenced by effects of late SECs on GSR that do not map to intensity changes. Its influence on the intensity layer should be inhibited here, e.g. by including an additional layer. But it remains an open question of study to identify these interacting components and stimulus evaluation checks.

The results of MSE are similar for the dynamic model and linear regression. But considering that estimation of affective states does not have to be a precision landing, dynamic intensity estimation is clearly superior to the linear regression estimate when inspecting Fig. 6.4. While bandwidth accuracy reflects this fact, MSE overemphasizes outliers and temporal lags. This confirms the argumentation that it is not a very suitable measure for the evaluation of emotion recognition in continuous space and time [203].

Further, comparison of the dynamic model with the performed linear regression by means of A_{bdw} at $b = .2$ suggests that b is chosen too large here, i.e. many samples fall into the range of 20% around ground truth without providing any qualitative insight on actual intensity levels.

In contrast, the failure of the dynamic model to outperform the Π-regressor for A_{bdw} at $b = .1$ marks the limitations of the current (simple) version of the model. While the dynamic model can provide rough information about the intensity ($b = .15$), it is not able to match intensity better than random within the fine ranges ($b = .1$). Here, other components and sequential evaluation checks need to be taken into account to improve results.

6.5 Conclusion

In this chapter, we designed and carried out an exemplary study to estimate emotion intensity from physiological signals using the model presented in Chapter 5. In this, the dynamic model performed significantly better than baselines. Although the onset of intensity change is well estimated, some late GSR responses are falsely related with intensity changes. Given the simple form of the model these limitations had to be expected and are subject to future work. This exemplary study was only concerned with a single location of the DNFs

over the emotion quality ϑ. It will be interesting to see how parameters change for different locations of the field.

The difference of our work presented in both Chap. 5 and Chap. 6 to the two most relevant existing works from emotion recognition are the following: Compared to the approach by Bailenson et al. [172], i.e. to estimate emotion quality and intensity separately using linear regression, our framework incorporates appraisal as well as emotion dynamics in both lateral (quality transitions) and temporal (intensity evolvement) dimension. In contrast to the non-predictive study by Grewe et al. [40], we measure intensity of subjective feeling explicitly during the presented user study instead of in two dimensional VA space. This is a more reliable measure, which further accounts for forms of emotion dynamics. The development of an input device reproducing the polar coordinates could be beneficial when recording the complete affective state at once.

6.5.1 Limitations and outlook

In order to first gain more knowledge on dynamic emotion recognition from physiological signals, controlled experiments need to be conducted. The current version of the model as well as results are limited to those cases, where it can be assumed by experimental design that no factor other than SECs affect the physiological signals. Loosening this constraint requires a substantially better understanding of underlying processes and is still a major hurdle to real world applications.

Measuring online labels in only one dimension improves the reliability of ground truth, but is still a viable concern. In general, variability of estimation success between subjects can be explained by the complexity of the task and the ability to assess ones own subjective feeling online. For future studies, testing interoceptive awareness of a subject before the experiment could be used as an additional measure of ground truth reliability [193].

The proposed framework opens a wide field of experiments and directions of study. It stands to reason to improve the proposed model in several ways, e.g. adding other physiological signals and extending the model to include interactions between input layers and additional layers, in order to overcome current limitations and improve estimation quality. Various experiments to verify and refine the presented model on a single emotion quality are conceivable. For example, what effect do different types and lengths of induction stimuli have on subjectively felt intensity and model parameters? Further, it would be interesting to compare intensity dynamics of different fixed emotion qualities: How well does the model fit and how do parameters vary? Another step in the future would be to consider experimental designs with two or more emotion qualities to measure and model intensity transitions between them. Ultimately, these results can be merged into the DNF over emotion quality ϑ. As a far outlook, it seems intriguing to look at and model competition of emotion qualities when several cues are presented, which DFT is generally capable of.

7 Conclusions and future directions

7.1 Summary

This thesis is concerned with static and dynamic methods for emotion recognition from physiological signals. The topic covers an interdisciplinary field at the cross section of psychology and engineering, which requires an understanding of psychological concepts as well as human physiological signals and their accessibility through measurements. This knowledge is necessary for the proper design of user studies to collect reliable data, which is still a big challenge in this research endeavor. The main contributions of this thesis to the field lie in the investigation of ways to improve existing static methods based on classical machine learning approaches as well as the development of a novel dynamic method for emotion recognition. A brief summary of the problems tackled together with the most important results is given in the following.

While many FS methods have been benchmarked on both synthetic and real-world datasets, a detailed investigation of various interaction types as attempted is yet missing. In Chapter 3, we investigated the capabilities of feature selection (FS) methods to detect and exploit interactions of features, as they are likely to occur in features from physiological signals. To gain a better understanding of the implications of interacting features, we performed a study on artificially created data for an academic example comparing a selection of state-of-the-art filter methods. While univariate methods can only consider features individually, we find that the ability of different multivariate methods to detect interactions of features depends on the kind of interaction at hand. For small sample sizes, however, our results suggest that univariate methods are more robust and tend to not overfit the data as much as multivariate methods. Thus, applying multiple FS methods simultaneously yields more stable results and allows for a better interpretability.

In emotion recognition from EEG, a vast amount of possible features and electrode locations have been proposed. But it is not clear, which features and electrodes are most suitable for emotion recognition. We propose to approach this issue by means of a systematic analysis of features using machine learning methods presented in Chapter 4. In this, an extensive survey of over 30 studies on proposed feature extraction methods is given, followed by the recording of a sufficiently large dataset containing EEG signals of eleven subject in five emotional states on which these features are compared. We carried out systematic analysis of features and electrode locations by applying multiple FS methods to identify most promising features and electrodes. While most findings overlap with previous studies on emotional brain activity, we suggest to promote some promising features from parietal and centro-parietal lobes. However, repeated experiments of the suggested procedure on other datasets will be necessary to reach a consensus on this topic.

In the second part of the thesis, we focus on the development of a novel dynamic model for emotion recognition. Although emotions are known to be dynamic processes, the consideration of dynamics has mostly been neglected in emotion recognition from physiological signals so far. In Chapter 5, we review the approach for emotion recognition with the aim of including findings from modern emotion theory, particularly appraisal models. A contribution to the field, which has not been attempted before, is made by the development of a dynamic gray-box model, which is capable of combining theoretical knowledge with data-driven parameter identification from experimental data. The proposed model is very general in that it allows to consider various types of emotion dynamics. Additionally, an in-detail version of the model for a specific experimental design is discussed.

We validate our proposed model in Chapter 6 by carrying out a user study which is concerned with the estimation of emotion intensity from physiological signals. In this, we present a new experimental design to record subjectively felt intensity levels in combination with physiological signals, here GSR. Since the estimation involves a continuous scale, we provide a discussion of the most relevant evaluation measures for this task in emotion recognition. The results from three subjects show a promising potential of the dynamic model. Since this approach is still in its infancy, a simple form of the model was chosen and the observed limitations had to be expected.

7.2 Outlook

Based on the topics considered in this thesis, a number of interesting research questions can be deduced from the presented results and are discussed in the following. Other individual aspects of future work are discussed at the end of each chapter.

Multivariate feature selection. In further improving the understanding and performance of feature selection techniques, several steps can be indicated. Although there already exists a large quantity of FS methods, the need to develop new algorithms is still substantial. In this, investigating in new evaluation criteria of subsets in combination with different search strategies seems promising. Especially randomized search strategies may help to find good solutions by avoiding local minima and at the same time lower computational costs of large problems.

Additionally, efforts in studies on artificial data may be extended by varying the degree of complexity of the FS task, which can be achieved by, for example, increasing the number of classes, the number of interacting features, and the number of irrelevant or redundant features compared to the study presented here. Further, the performance of each method could be evaluated for combinations of interaction types.

EEG datasets. As mentioned above, application of the proposed systematic analysis for feature selection on other datasets are necessary to reach a consensus on this topic. Besides possibly extending the selection of FS methods used for the analysis, the characteristics of future datasets should be broadly based. In this, different design factors should be accounted for by varying, for example, the induction methods and the selection of emotional states as

well as by extending the datasets over several sessions on different days. In approaching the design of new databases in a similarly systematic fashion, we believe that meaningful conclusions on many questions can be drawn.

Emotion dynamics. A better understanding of the emotion process should be a major focus of future research. There, both experiments and methodological developments need to incorporate the implications involved in the dynamic nature of emotions. This includes work that investigates in how emotions unfold, how emotional states transfer, and how different types of emotion dynamics interact with each other. Ultimately, a more accurate model of the emotion process will make it easier to select and develop appropriate methods for emotion recognition.

Real world applications. As an outlook on real-world applications, it might not be obvious how findings from the lab can be transfered to the real world. One way of achieving this step is to loosen experimental conditions incrementally, e.g. by introducing secondary tasks when recoding emotions, and test whether findings from very controlled experiments still hold. It is also entailed in this aspect that we encourage the development of novel and more reliable methods for emotion induction, for example, by taking advantage of the great advances in media technology and social networks.

A final remark concerns the interdisciplinarity of this research field, which makes collaboration between psychologists and engineers key. Projects and workshops of this kind should be fostered to facilitate entry to the topic and keep people hooked to its compelling challenges, especially when they come from one field with no prior experience in the other.

A Appendix

A.1 Tools for user studies

Several tools to aid the experiments carried out have been developed. With the intention of passing on useful knowledge to the interested reader, some key facts on these tools are provided here.

A.1.1 IAPS picture selector

The IAPS dataset consists of more than 1100 pictures, which makes hand selection either tedious or prone to human-caused bias when selecting stimuli for emotion induction. Therefore, a software tool was developed that can import the ratings from text files accompanying the documentation of the dataset and automatically filters out appropriate pictures corresponding to a range of VAD values and allowed variance provided by the user. Additionally, the maximum number of pictures can be specified. In this case, most "appropriate" is defined as the pictures that fulfill the search criteria with least distance from the mean VAD value and lowest variance in VAD ratings, where the weighting of both criteria is defined by the user. The list of pictures resulting from a query can automatically be copied into a new folder, which is later accessed by the induction software tool (see next section). A screen shot of the GUI is shown in Fig. A.1.
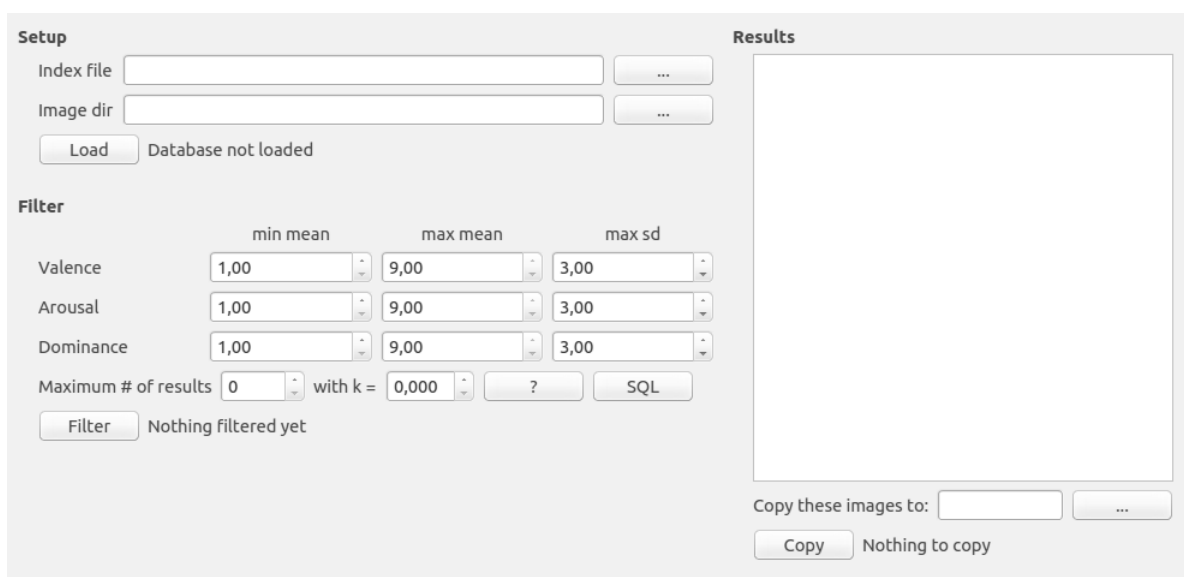

Figure A.1: IAPS filter GUI to select appropriate stimuli automatically.

A.1.2 Induction software tool

When carrying out experiments involving human subjects, it is crucial that the experimental conditions are controlled for as consistently as possible across all subjects, e.g. by giving exactly the same instructions and adhere to the protocol. To minimize effects of varying conditions, an induction software tool was developed in order to automate the experimental protocol as well as randomizing experimental conditions. In this, it automatically guides the subject through the experimental process without the need for an instructor to interfere. Further, the software collected all relevant data from the subject during the experiment, including user ID, which is a required input right after launching the program, and self-reports of the emotional experience by means of a SAM test. The software is configured via an XML file and allows for user-definition of the following options:

- Display options: among graphical setting like font-size and full-screen mode, most important settings regard the number of emotions (a set of emotions refers to passing through each emotion condition once) and number of pictures per emotion, the display time of each picture, and black screen times before or after the pictures. Additionally, neutral pictures in between the trials can be configured as well as several screen messages like instructions to the subjects.

- Log-file: the user ID, the order of emotions and filenames of the pictures shown as well as the SAM test results are stored in a local file to permit a detailed post-hoc analysis of induction quality.

- UDP settings: hosts and ports for synchronization signals via UDP are defined in the configuration file. The user ID is sent at the beginning of each set of emotions; at the beginning and end of each recording trial, an emotion identifier is transferred which facilitates automated segmentation of the data; the SAM results are sent after completion of each test.

- Pictures: the folders in which selected IAPS pictures for induction are stored are indicated by the user and automatically loaded by the software tool, i.e. it finds and subsequently uses all graphic files contained in the specified folder for one emotion. Within one set of emotions, the order of emotions is randomized. Further, the induction software tool randomly picks and displays the specified number of pictures for each trial and remembers which pictures have already been shown to the current user ID so that each picture is only shown once.

The induction software tool has been successfully used to carry out both experiments reported in Chapter 4 and Chapter 6.

A.1.3 Emotion intensity slider

The custom-made emotion slider device (shown in Fig. A.2a) used to dynamically record emotion intensity is based on a linear potentiometer with a travel of 11 cm and can be fixed to the arm rest of the subject's chair. It is connected to the Arduino UNO interface (shown in Fig. A.2b) and is powered by its 5 V power supply. The current slider position is determined using the analog input of the Arduino UNO.
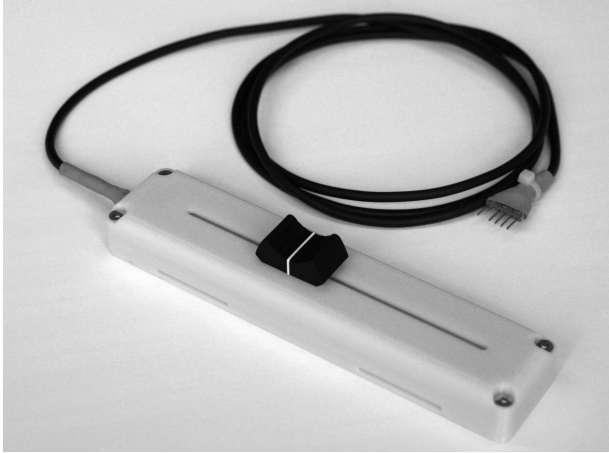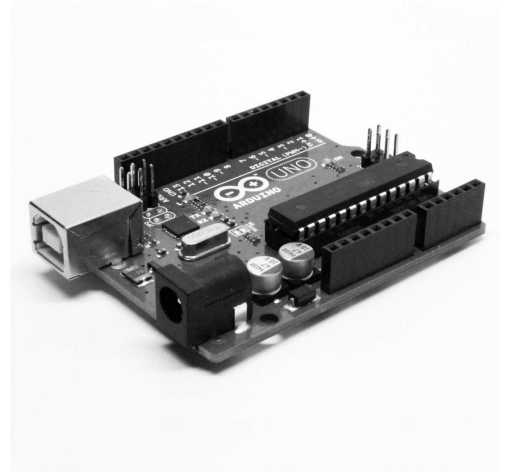
(a)   (b)

Figure A.2: (a) The slider device together with (b) the Arduino interface [194] was used in the experiment in Chapter 6.

A.2 Model parameters of emotion intensity study

When identifying model parameters using optimization, lower and upper bounds of the parameters $\mathbf{p}$ listed in Section 5.4.4 are defined as given in Table A.1. Tables A.2, A.3, and A.4 list the parameter values that were identified in the study in Chapter 6 for each test trial from leave-one-out cross validation.

Tabular A.1: Lower and upper bounds ($\mathbf{p}_{min/max}$) for parameters $\mathbf{p}$ of dynamic model for a single emotion quality.

$\mathbf{p}$	τ_u	τ_i	τ_m	h_u	h_i	h_m	β_u	β_i
$\mathbf{p}_{min}$	0	0	0	-20	-20	-20	1	1
$\mathbf{p}_{max}$	50	50	50	0	0	0	200	200

$\mathbf{p}$	β_m	c_{uu}	c_{ii}	c_{mm}	c_{iu}	c_{im}	c_{mu}	c_{mi}
$\mathbf{p}_{min}$	1	-10	-10	-10	-10	-10	-10	-10
$\mathbf{p}_{max}$	200	10	10	10	10	10	10	10

Tabular A.2: Values of parameters **p** of dynamic model for a single emotion quality for each trial of Subject 1.

Parameter	1	2	3	4	5	6	7
τ_u	7.01	3.20	7.01	4.65	3.02	7.01	3.34
τ_i	31.93	29.97	31.93	30.27	29.94	31.93	29.96
τ_m	20.07	20.00	20.07	20.02	20.00	20.07	20.00
h_u	−1.38	−4.94	−1.38	−2.97	−4.93	−1.38	−4.75
h_i	−0.64	−0.54	−0.64	−0.51	−0.53	−0.64	−0.53
h_m	0.00	−0.05	0.00	−0.11	0.00	0.00	−0.04
β_u	100.00	100.00	100.00	100.00	100.00	100.00	100.00
β_i	4.07	5.08	4.07	5.30	5.04	4.07	5.12
β_m	49.99	50.00	49.99	49.98	50.00	49.99	50.00
c_{uu}	2.33	0.06	2.33	2.26	0.01	2.33	0.18
c_{ii}	−1.41	−1.04	−1.41	−0.89	−1.41	−1.41	−1.16
c_{mm}	2.04	1.83	2.04	1.23	1.69	2.04	1.94
c_{iu}	−0.52	0.57	−0.52	0.41	0.34	−0.52	0.58
c_{im}	−2.37	−2.10	−2.37	−1.94	−2.16	−2.37	−2.09
c_{mu}	1.37	1.10	1.37	1.17	0.99	1.37	1.12
c_{mi}	−0.13	0.73	−0.13	0.42	0.69	−0.13	0.69

Tabular A.3: Values of parameters **p** of dynamic model for a single emotion quality for each trial of Subject 2.

Parameter	1	2	3	4	5	6	7
τ_u	1.95	3.42	2.92	3.06	3.14	2.92	2.92
τ_i	31.41	29.95	29.96	30.34	30.16	29.96	29.96
τ_m	19.80	19.99	19.99	20.00	20.00	19.99	19.99
h_u	−2.43	−5.01	−5.25	−4.69	−5.69	−5.25	−5.25
h_i	−0.88	−0.45	−0.49	−0.61	−0.42	−0.49	−0.49
h_m	−1.69	−0.04	−0.04	−0.34	0.00	−0.04	−0.04
β_u	100.00	100.00	100.00	100.00	100.00	100.00	100.00
β_i	4.19	5.26	5.24	5.02	5.49	5.24	5.24
β_m	50.01	50.00	50.00	50.00	50.00	50.00	50.00
c_{uu}	0.68	0.02	−0.57	0.02	1.02	−0.57	−0.57
c_{ii}	4.11	−2.16	−1.66	0.16	−2.51	−1.66	−1.66
c_{mm}	3.69	2.44	2.25	2.97	2.97	2.25	2.25
c_{iu}	−4.22	0.08	0.37	−2.33	0.29	0.37	0.37
c_{im}	0.20	−2.29	−2.29	−2.03	−2.27	−2.29	−2.29
c_{mu}	3.85	1.07	0.95	1.38	1.19	0.95	0.95
c_{mi}	1.95	0.48	0.49	1.07	0.27	0.49	0.49

Tabular A.4: Values of parameters **p** of dynamic model for a single emotion quality for each trial of Subject 3.

Parameter	1	2	3	4	5	6	7
τ_u	2.73	3.27	3.21	3.46	2.73	2.91	2.73
τ_i	29.93	29.90	29.91	29.89	29.93	29.88	29.93
τ_m	20.00	20.00	20.00	19.99	20.00	19.99	20.00
h_u	−4.90	−5.01	−5.04	−5.00	−4.90	−5.18	−4.90
h_i	−0.52	−0.45	−0.45	−0.43	−0.52	−0.48	−0.52
h_m	0.00	−0.97	−1.03	−0.88	0.00	−0.00	0.00
β_u	100.00	100.00	100.00	100.00	100.00	100.00	100.00
β_i	5.33	5.84	5.82	6.16	5.33	5.25	5.33
β_m	50.00	50.00	50.00	50.00	50.00	50.00	50.00
c_{uu}	0.01	0.02	0.02	0.06	0.01	−0.13	0.01
c_{ii}	−1.27	−1.04	−1.05	−1.01	−1.27	−1.88	−1.27
c_{mm}	2.52	3.04	3.08	3.22	2.52	2.33	2.52
c_{iu}	0.69	1.34	1.37	1.31	0.69	0.25	0.69
c_{im}	−2.06	−1.82	−1.81	−1.81	−2.06	−2.25	−2.06
c_{mu}	1.36	1.45	1.41	1.64	1.36	1.10	1.36
c_{mi}	0.61	0.76	0.74	0.91	0.61	0.51	0.61

B Bibliography

[1] A. Brooks and R. Arkin, "Behavioral overlays for non-verbal communication expression on a humanoid robot," *Autonomous Robots*, vol. 22, pp. 55–74, 2007.

[2] C. Breazeal, "Emotion and sociable humanoid robots," *Int. Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 119–155, 2003.

[3] C. Bartneck, J. Reichenbach, and A. Breemen, "In your face, robot! The influence of a character's embodiment on how users perceive its emotional expressions," in *Proc. of Design and Emotion*, Ankara, 2004.

[4] C. Lisetti and F. Nasoz, "Using noninvasive wearable computers to recognize human emotions from physiological signals," *EURASIP Journal on Advances in Signal Processing*, vol. 11, pp. 1672–1687, 2004.

[5] R. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175–1191, 2001.

[6] A. Haag, S. Goronzy, P. Schaich, and J. Williams, "Emotion recognition using bio-sensors: First steps towards an automatic system," in *Affective Dialogue Systems*, 2004, pp. 36–48.

[7] C. Mühl, B. Allison, A. Nijholt, and G. Chanel, "A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges," *Brain-Computer Interfaces*, vol. 1, no. 2, pp. 66–84, 2014.

[8] S. I. Hjelm, "Research and design: The making of Brainball," *Interactions*, vol. 10, no. 1, pp. 26–34, 2003.

[9] L. George, F. Lotte, R. Abad, and A. Lecuyer, "Using scalp electrical biosignals to control an object by concentration and relaxation tasks: design and evaluation," English, in *Proc. of the 33th Int. Conf. of the IEEE Engineering in Medicine and Biology Society.*, 2011, pp. 6299–302.

[10] R. Picard, "Affective computing," MIT Media Laboratory, Cambridge, MA, MIT Technical Report #321, 1995.

[11] T. Zander and C. Kothe, "Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general.," *Journal of neural engineering*, vol. 8, no. 2, p. 025 005, 2011.

[12] N. Frijda, *The emotions*. Cambridge University Press, 1986.

[13] E. T. Rolls, *The Brain and Emotion*. Oxford University Press, 1999.

[14] P. Kleinginna and A. Kleinginna, "A categorized list of motivation definitions, with a suggestion for a consensual definition," *Motivation and Emotion*, vol. 5, no. 4, pp. 263–291, 1981.

[15] C. Smith and R. Lazarus, "Emotion and adaptation," in *Handbook of Personality: Theory and Research*, L. Pervin, Ed., NewYork: Guilford, 1990, pp. 609–637.

[16] R. Ketai, "Affect, mood, emotion, and feeling: semantic considerations," *American Journal of Psychiatry*, vol. 132, no. 11, pp. 1215–1217, 1975.

[17] K. Scherer, "What are emotions? And how can they be measured?" *Social Science Information*, vol. 44, no. 4, pp. 695–729, 2005.

[18] R. Lazarus, *Emotion & Adaption*. Oxford University Press, 1991.

[19] R. Cowie, G. McKeown, and E. Douglas-Cowie, "Tracing emotion," *Int. Journal of Synthetic Emotions*, vol. 3, no. 1, pp. 1–17, 2012.

[20] P. Ekman and R. Davidson, "Moods, emotions and traits," in *The Nature of Emotion*, P. Ekman and R. Davidson, Eds., Oxford University Press, 1994, pp. 56–8.

[21] P. Ekman and W. Friesen, "Constants across cultures in the face and emotion.," *Journal of Personality and Social Psychology*, vol. 17, pp. 124–129, 1971.

[22] P. Ekman, "Expression and the nature of emotion," in *Approaches to emotion*, K. Scherer and P. Ekman, Eds., Hillsdale, NJ: Lawrence Erlbaum, 1984, pp. 319–343.

[23] P. Ekman and W. Friesen, "A new pan-cultural facial expression of emotion," *Motivation and Emotion*, vol. 10, no. 2, pp. 159–168, 1986.

[24] A. Ortony and T. Turner, "What's basic about basic emotions?" *Psychological review*, vol. 97, no. 3, pp. 315–331, 1990.

[25] J. Tracy and D. Randles, "Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt," *Emotion Review*, vol. 3, no. 4, pp. 397–405, 2011.

[26] J. Russell and A. Mehrabian, "Evidence for a three-factor theory of emotions," *Journal of Research in Personality*, vol. 11, pp. 273–94, 1977.

[27] J. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[28] J. Broekens, "In defense of dominance: PAD usage in computational representations of affect," *Int. Journal of Synthetic Emotions (IJSE)*, vol. 3, no. 1, pp. 33–42, 2012.

[29] R. Plutchik, "The nature of emotions," *American Scientist*, vol. 89, pp. 344–350, 2001.

[30] M. Yik, J. Russell, and J. Steiger, "A 12-Point circumplex structure of core affect," *Emotion*, vol. 11, no. 4, pp. 705–31, 2011.

[31] A. Moors, P. Ellsworth, K. Scherer, and N. Frijda, "Appraisal theories of emotion: State of the art and future development," *Emotion Review*, vol. 5, no. 2, pp. 119–124, 2013.

[32] K. Scherer, "Emotions are emergent processes: they require a dynamic computational architecture.," *Phil. Trans. of the Royal Society, Series B*, vol. 364, pp. 3459–74, 2009.

[33] S. Marsella, J. Gratch, and P. Petta, "Computational models of emotion," in *A Blueprint for an Affectively Competent Agent*, K. Scherer, T. Bänziger, and E. Roesch, Eds., Oxford University Press, 2010, pp. 21–46.

[34] M. Mortillaro, B. Meuleman, and K. Scherer, "Advocating a componential appraisal model to guide emotion recognition," *Int. Journal of Synthetic Emotions*, vol. 3, no. 1, pp. 18–32, 2012.

[35] R. Reisenzein, "Pleasure-arousal theory and the intensity of emotions.," *Journal of Personality and Social Psychology*, vol. 67, no. 3, pp. 525–539, 1994.

[36] E. Harmon-Jones, D. Amodio, and L. Zinner, "Social psychological methods of emotion elicitation," in *Handbook of Emotion Elicitation and Assessment*, J. Coan and J. Allen, Eds., Oxford University Press, 2007, pp. 91–105.

[37] C. De Dreu, M. Baas, and B. Nijstad, "Hedonic tone and activation level in the mood-creativity link: toward a dual pathway to creativity model.," *Journal of Personality and Social Psychology*, vol. 94, no. 5, pp. 739–756, 2008.

[38] C. Kothe, S. Makeig, and J. Onton, "Emotion recognition from EEG during self-paced emotional imagery," in *IEEE Int. Conf. on Affective Computing and Intelligent Interaction*, 2013, pp. 855–858.

[39] M. Bradley and P. Lang, "The International Affective Digitized Sounds: Affective Ratings of Sounds and Instruction Manual," University of Florida, Gainesville, FL, Tech. Rep., 2007, pp. 29–46.

[40] O. Grewe, F. Nagel, R. Kopiez, and E. Altenmüller, "Emotions over time: synchronicity and development of subjective, physiological, and facial affective reactions to music.," *Emotion*, vol. 7, no. 4, pp. 774–88, 2007.

[41] P. Lang, M. Bradley, and B. Cuthbert, "International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual," University of Florida, Gainesville, FL, Tech. Rep., 2005, pp. 1–56.

[42] S. Koelstra, C. Mühl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis using physiological signals," *IEEE Trans. on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[43] Y. Baveye, E. Dellandr, L. Chen, and C. Chamaret, "A Large Video Database for Computational Models of Induced Emotion," 2013.

[44] H. Gunes and M. Pantic, "Automatic, dimensional and continuous emotion recognition," *Int. Journal of Synthetic Emotions*, vol. 1, no. 1, pp. 68–99, 2010.

[45] M Bradley and P. J. Lang, "Measuring emotion: The Self-Assessment Manikin and the semantic differential," *Journal of Behavioral Theory and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[46] G. Caridakis, K. Karpouzis, and S. Kollias, "User and context adaptive neural networks for emotion recognition," *Neurocomputing*, vol. 71, no. 13-15, pp. 2553–2562, 2008.

[47] *http://www.gtec.at*. g.tec Medical Engineering, company website, accessed March, 2015.

[48] A. Nijholt and D. Tan, "Brain-computer interfacing for intelligent systems," *IEEE Intelligent Systems*, vol. 23, no. 3, pp. 72–79, 2008.

[49] A. Mognon, J. Jovicich, L. Bruzzone, and M. Buiatti, "ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features.," *Psychophysiology*, vol. 48, pp. 229–240, 2010.

[50] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis.," *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9–21, 2004.

[51] G. Valenza, A. Lanatà, and E. Scilingo, "The role of nonlinear dynamics in affective valence and arousal recognition," *IEEE Transactions on Affective Computing*, vol. 3, no. 2, pp. 237–249, 2012.

[52] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning*, 2nd Editio. Springer, 2009.

[53] A. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, pp. 199–222, 2004.

[54] D. Hunter, H. Yu, S. Member, M. Pukish, J. Kolbusz, and B. Wilamowski, "Selection of proper Neural Network sizes and architectures - A comparative study," *IEEE Transactions on Inductrial Informatics*, vol. 8, no. 2, pp. 228–240, 2012.

[55] K. Murphy, *Machine Learning - a probabilistic perspective*. The MIT Press, 2012.

[56] J. Wagner, E André, and F. Jung, "Smart sensor integration: A framework for multimodal emotion recognition in real-time," in *IEEE Int. Conf. on Affective Computing and Intelligent Interaction*, 2009, pp. 1–8.

[57] S. Gilroy, M. Cavazza, M. Niiranen, E. André, T. Vogt, J. Urbain, M. Benayoun, H. Seichter, and M. Billinghurst, "PAD-based multimodal affective fusion," in *IEEE Int. Conf. on Affective Computing and Intelligent Interaction*, Ieee, 2009, pp. 1–8.

[58] E. van den Broek, M. Schut, J. Westerink, and K. Tuinenbreijer, "Unobtrusive sensing of emotions (USE)," *Journal of Ambient Intelligence and Smart Environments*, vol. 1, pp. 287–299, 2009.

[59] E. Leon, G. Clarke, V. Callaghan, and F. Sepulveda, "A user-independent real-time emotion recognition system for software agents in domestic environments," *Engineering Applications of Artificial Intelligence*, vol. 20, no. 3, pp. 337–345, 2007.

[60] B. Grundlehner, L. Brown, J. Penders, and B. Gyselinckx, "The design and analysis of a real-time, continuous arousal monitor," in *IEEE Int. Workshop on Wearable and Implantable Body Sensor Networks*, Ieee, 2009, pp. 156–161.

[61] R. Sitaram, S. Lee, S. Ruiz, M. Rana, R. Veit, and N. Birbaumer, "Real-time support vector classification and feedback of multiple emotional brain states.," *NeuroImage*, vol. 56, no. 2, pp. 753–65, 2011.

[62] E. Konstantinidis, C. Frantzidis, C. Pappas, and P. Bamidis, "Real time emotion aware applications: a case study employing emotion evocative pictures and neuro-physiological sensing enhanced by graphic processor units.," *Computer Methods and Programs in Biomedicine*, vol. 107, no. 1, pp. 16–27, 2012.

[63] Y. Liu and O. Sourina, "Real-time fractal-based valence level recognition from EEG," *Trans. on Computational Science XVIII*, vol. 7848, pp. 101–120, 2013.

[64] K. Schaaff and M. Adam, "Measuring emotional arousal for online applications: Evaluation of ultra-short term heart rate variability measures," in *IEEE Int. Conf. on Affective Computing and Intelligent Interaction*, 2013, pp. 362–368.

[65] J. Healey, "Affect detection in the real world: Recording and processing physiological signals," in *IEEE Int. Conf. Affective Computing and Intelligent Interaction*, 2009, pp. 1–6.

[66] M. Lichman, "UCI Machine Learning Repository," http://archive.ics.uci.edu/ml, University of California, Irvine, School of Information and Computer Sciences, 2013.

[67] A. Jakulin and I. Bratko, "Analyzing attribute dependencies," *LNCS*, vol. 2838, pp. 229–240, 2003.

[68] Z. Zhao and H. Liu, "Searching for interacting features in subset selection," *Intelligent Data Analysis*, vol. 13, pp. 207–228, 2007.

[69] Z. Zhu, Y.-S. Ong, and J. Zurada, "Identification of full and partial class relevant genes.," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 7, no. 2, pp. 263–77, 2010.

[70] S. Thrun, J. Bala, E. Bloedom, I. Bratko, B. Cestnik, J. Cheng, K. De Jong, S. Dzeroski, D. Fisher, S. Fahlman, R. Hamann, K. Kaufman, S. Keller, I. Kononenko, J. Kreuziger, R. Michalski, T. Mitchell, P. Pachowicz, Y. Reich, H. Vafaie, W. van de Welde, W. Wenzel, J. Wnek, and J. Zhang, "The MONK's problems: A performance comparison of different learning algorithms," Carnegie Mellon University, Technical Report CMU-CS-91-197, 1991.

[71] L. Belanche and F. González, "Review and evaluation of feature selection algorithms in synthetic problems," *CoRR*, vol. abs/1101.2, 2011. arXiv: 1101.2330.

[72] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, "On the behavior of feature selection methods dealing with noise and relevance over synthetic scenarios," in *IEEE Int. Conf. on Neural Networks*, 2011, pp. 1530–1537.

[73] R. Bellman, *Adaptive control processes - A guided tour*, Princeton, NJ, 1961.

[74] I. Jolliffe, *Principal component analysis*, 2nd Ed. Springer Series in Statistics, 2002.

[75] R. Shepard, "Multidimensional Scaling, Tree-Fitting, and Clustering," *Science*, vol. 210, no. 4468, pp. 390–398, 1980.

[76] M. Dash and H. Liu, "Feature selection for classification," *Intelligent Data Analysis*, vol. 1, no. 3, pp. 131–156, 1997.

[77] A. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial Intelligence*, vol. 97, pp. 245–271, 1997.

[78] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *The Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[79] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics.," *Bioinformatics*, vol. 23, no. 19, pp. 2507–17, 2007.

[80] H. Liu and H. Motoda, *Computational Methods of Feature Selection*. Chapman & Hall/CRC, 2008.

[81] P. Somol, J. Novovicová, and P. Pudil, "Efficient feature subset selection and subset size optimization," in *Pattern Recognition Recent Advances*, A. Herout, Ed., InTech, 2010, pp. 1–24.

[82] C. Lazar, J. Taminau, S. Meganck, D. Steenhoff, A. Coletta, C. Molter, V. de Schaetzen, R. Duque, H. Bersini, and A. Nowé, "A survey on filter techniques for feature selection in gene expression microarray analysis.," *IEEE/ACM Trans. on Computational Biology and Bioinfromatics*, vol. 9, no. 4, pp. 1106–19, 2012.

[83] K. Torkkola, "Feature Extraction by non-parametric mutual information maximization," *Journal of Machine Learning Research*, vol. 3, pp. 1415–1438, 2003.

[84] L. Deng, J. Pei, J. Ma, and D. Lee, "A rank sum test method for informative gene discovery," in *ACM Int. Conf. Knowledge Discovery and Data Mining*, 2004, pp. 410–419.

[85] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Machine Learning*, vol. 53, pp. 23–69, 2003.

[86] M. Karg, "Pattern Recognition Algorithms for Gait Analysis with Application to Affective Computing," PhD thesis, Technische Universität München, 2012.

[87] S. Wang, D. Li, Y. Wei, and H. Li, "A feature selection method based on fisher's discriminant ratio for text sentiment classification," *LNCS Web Information Systems and Mining*, vol. 5854, pp. 88–97, 2009.

[88] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. Routledge, 1988.

[89] M. Hall, "Correlation-based feature selection for machine learning," PhD thesis, University of Waikato, 1999.

[90] M. Deriche and A. Al-Ani, "A new algorithm for EEG feature selection using mutual information," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, 2001, pp. 1057–1060.

[91] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data.," in *Proc. of the Computational Systems Bioinformatics*, vol. 3, 2003.

[92] D. Koller and M. Sahami, "Toward optimal feature selection," in *Int. Conf. on Machine Learning*, 1996, pp. 284–292.

[93] H. Mamitsuka, "Selecting features in microarray classification using ROC curves," *Pattern Recognition*, vol. 39, no. 12, pp. 2393–2404, 2006.

[94] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data.," *Journal of Bioinformatics and Computational Biology*, vol. 3, no. 2, pp. 185–205, 2005.

[95] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–38, 2005.

[96] C. Huberty and S. Olejnik, *Applied MANOVA and Discriminant Analysis*, 2nd Ed. Wiley, 2006.

[97] A. El Ouardighi, A. El Akadi, and D. Aboutajdine, "Feature selection on supervised classification using Wilk's Lambda statistic," in *Int. Symp. on Computational Intelligence and Intelligent Infarmatics (ISCIII)*, 2007, pp. 51–55.

[98] A. Jakulin, "Machine learning based on attribute interactions," PhD thesis, University of Ljubljana, 2005.

[99] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, "A review of feature selection methods on synthetic data," *Knowledge and Information Systems*, vol. 34, no. 3, pp. 483–519, 2013.

[100] I. Witten, E. Frank, and M. Hall, *Data Mining*, 3rd Ed. Elsevier, 2011.

[101] L. Kuncheva, "A stability index for feature selection," in *Proc. of the 25th Int. Conf. on Artificial Intelligence and Applications*, 2007, pp. 390–395.

[102] A. Kalousis, J. Prados, and M. Hilario, "Stability of feature selection algorithms: a study on high-dimensional spaces," *Knowledge and Information Systems*, vol. 12, no. 1, pp. 95–116, 2007.

[103] S. Student and K. Fujarewicz, "Stable feature selection and classification algorithms for multiclass microarray data.," *Biology Direct*, vol. 7, no. 33, pp. 1–20, 2012.

[104] C. Mühl, A.-M. Brouwer, N. van Wouwe, E. van den Broek, F. Nijboer, and D. Heylen, "Modality-specific affective responses and their implications for affective BCI," in *Int. Conf. on Brain-Computer Interfaces*, 2011, pp. 4–7.

[105] T. Musha, Y. Terasaki, H. Haque, and G. Ivamitsky, "Feature extraction from EEGs associated with emotions," *Artificial Life and Robotics*, vol. 1, no. 1, pp. 15–19, 1997.

[106] M.-K. Kim, M. Kim, E. Oh, and S.-P. Kim, "A review on the computational methods for emotional state estimation from the human EEG.," *Computational and Mathematical Methods in Medicine*, no. 573734, pp. 1–13, 2013.

[107] J. Cacioppo, "Feelings and emotions: roles for electrophysiological markers.," *Biological Psychology*, vol. 67, no. 1-2, pp. 235–43, 2004.

[108] S. Sanei and J. Chambers, *EEG Signal Processing*. Wiley, 2007.

[109] K. Schaaff and T. Schultz, "Towards emotion recognition from electroencephalographic signals," in *Int. Conf. on Affective Computing and Intelligent Interaction (ACII)*, 2009, pp. 175–180.

[110] S. Hadjidimitriou and L. Hadjileontiadis, "Toward an EEG-based recognition of music liking using time-frequency analysis.," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 12, pp. 3498–510, 2012.

[111] P. Petrantonakis and L. Hadjileontiadis, "Emotion recognition from EEG using higher order crossings," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 186–197, 2010.

[112] K. Takahashi, "Remarks on emotion recognition from multi-modal bio-potential signals," in *Int.Conf. on Industrial Technology (ICIT)*, vol. 3, 2004, pp. 1138–1143.

[113] M. Murugappan, M. Rizon, S. Yaacob, I. Zunaidi, and D. Hazry, "EEG feature extraction for classifying emotions using FCM and FKM," *Int. Journal of Computer and Communications*, vol. 1, no. 2, pp. 21–25, 2007.

[114] X. Wang, D. Nie, and B. Lu, "EEG-based emotion recognition using frequency domain features and support vector machines," in *Int. Conf. on Neural Information Processing (ICONIP)*, 2011, pp. 734–743.

[115] K. Ansari-asl, G. Chanel, and T. Pun, "A channel selection method for EEG classification in emotion assessment based on synchronization likelihood," in *15th European Signal Processing Conference (EUSIPCO)*, 2007, pp. 1241–1245.

[116] C. Frantzidis, C. Bratsas, C. Papadelis, E. Konstantinidis, C. Pappas, and P. Bamidis, "Toward emotion aware computing: An integrated approach using multichannel neurophysiological recordings and affective visual stimuli," *IEEE Trans. on Information Technology in Biomedicine*, vol. 14, no. 3, pp. 589–597, 2010.

[117] Z. Lan, O. Sourina, L. Wang, and Y. Liu, "Stability of features in real-time EEG-based emotion recognition algorithm," in *IEEE Int. Conf. on Cyberworlds*, 2014, pp. 137–44.

[118] M. Murugappan, R. Nagarajan, and S. Yaacob, "Classification of human emotion from EEG using discrete wavelet transform," *Journal of Biomedical Science and Engineering*, vol. 03, no. 04, pp. 390–396, 2010.

[119] E. Kroupi, A. Yazdani, and T. Ebrahimi, "EEG correlates of different emotional states elicited during watching music videos," in *Int. Conf. on Affective Computing and Intelligent Interaction*, 2011, pp. 457–466.

[120] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalography and Clinical Neurophysiology*, vol. 29, no. 3, pp. 306–310, 1970.

[121] R. Horlings, D. Datcu, and L. Rothkrantz, "Emotion recognition using brain activity," in *Int. Conf. on Computer Systems and Technologies (CompSysTech)*, 2008, pp. II.1–1–6.

[122] J. Hausdorff, A. Lertratanakul, M. Cudkowicz, A. Peterson, D. Kaliton, and A. Gold-berger, "Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis," *Journal of Applied Physiology*, vol. 88, pp. 2045–2053, 2000.

[123] R. Khosrowabadi and A. Rahman, "Classification of EEG correlates on emotion using features from Gaussian mixtures of EEG spectrogram," in *Proc. of the 3rd Int. Conf. on Information and Communication Technology for the Moslem World (ICT4M)*, 2010, E102–E107.

[124] O. Sourina and Y. Liu, "A fractal-based algorithm of emotion recognition from EEG using arousal-valence model.," in *BIOSIGNALS*, 2011, pp. 209–214.

[125] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica D: Nonlinear Phenomena*, vol. 31, no. 2, pp. 277–283, 1988.

[126] P. Petrantonakis and L. Hadjileontiadis, "Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis," *IEEE Transactions on Affective Computing*, vol. 1, no. 2, pp. 81–97, 2010.

[127] B. Reuderink, C. Mühl, and M. Poel, "Valence, arousal and dominance in the EEG during game play," *Int. Journal of Autonomous and Adaptive Communications Systems*, vol. 6, no. 1, pp. 45–62, 2013.

[128] L. Brown, B. Grundlehner, and J. Penders, "Towards wireless emotional valence detection from EEG.," in *IEEE Int. Conf. of the Engineering in Medicine and Biology Society*, 2011, pp. 2188–91.

[129] G. Chanel, K. Ansari-asl, and T. Pun, "Valence-arousal evaluation using physiological signals in an emotion recall paradigm," in *IEEE Int. Conf. on Systems, Man and Cybernetics (ISIC)*, 2007, pp. 2662–2667.

[130] A. Conneau and S. Essid, "Assessment of new spectral features for EEG-based emotion recognition," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, 2014, pp. 4698–4702.

[131] S. Jirayucharoensak, P.-N. Setha, and P. Israsena, "EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation," *The Scientific World Journal*, no. 627892, pp. 1–10, 2014.

[132] Y.-Y. Lee and S. Hsieh, "Classifying different emotional states by means of EEG-based functional connectivity patterns," *PLoS ONE*, vol. 9, no. 4, e95415, 2014.

[133] Y. Liu and O. Sourina, "EEG-based dominance level recognition for emotion-enabled interaction," in *IEEE Int. Conf. on Multimedia and Expo*, 2012, pp. 1039–1044.

[134] D. Nie, X.-W. Wang, L.-C. Shi, and B.-L. Lu, "EEG-based emotion recognition during watching movies," in *IEEE Int. Conf. on Neural Engineering*, 2011, pp. 667–670.

[135] M. Li and B.-L. Lu, "Emotion classification based on gamma-band EEG," in *IEEE Int. Conf. of Engineering in Medicine and Biology Society*, vol. 1, 2009, pp. 1323–26.

[136] D. Bos, "EEG-based emotion recognition," *Emotion*, vol. 57, no. 7, pp. 1798–806, 2006.

[137] G. Lee, M. Kwon, S. Kavuri Sri, and M. Lee, "Emotion recognition based on 3D fuzzy visual and EEG features in movie clips," *Neurocomputing*, vol. 144, pp. 560–568, 2014.

[138] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for EEG-based emotion classification," in *IEEE/EMBS Int.Conf. on Neural Engineering*, 2013, pp. 81–84.

[139] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, and J.-H. Chen, "EEG-based emotion recognition in music listening.," *IEEE Trans. on Biomedical Engineering*, vol. 57, no. 7, pp. 1798–806, 2010.

[140] V. Rozgić, S. Vitaladevuni, and R. Prasad, "Robust EEG emotion classification using segment level decision fusion," in *Int. Conf. on Acustics, Sound, and Signal Processing (ICASSP)*, 2013, pp. 1286–1290.

[141] M. Soleymani, S. Koelstra, I. Patras, and T. Pun, "Continuous emotion detection in response to music videos," in *IEEE Int. Conf. on Automatic Face & Gesture Recognition*, 2011, pp. 803–8.

[142] S. Hosseini, M. Khalilzadeh, M. Naghibi-Sistani, and V. Niazmand, "Higher order spectra analysis of EEG signals in emotional stress states," in *IEEE Int. Conf. on Information Technology and Computer Science*, 2010, pp. 60–63.

[143] A. Swami, J. Mendel, and C. Nikias, "Higher-Order Spectra Analysis (HOSA) Toolbox, version 2.0.3," *available at www.mathworks.com/matlabcentral/fileexchange/3013/*, 2000.

[144] M. Akin, "Comparison of wavelet transform and FFT methods in the analysis of EEG signals.," *Journal of Medical Systems*, vol. 26, no. 3, pp. 241–7, 2002.

[145] C. Parameswariah and M. Cox, "Frequency characteristics of wavelets," *IEEE Trans. on Power Delivery*, vol. 17, no. 3, pp. 800–4, 2002.

[146] M. Murugappan, M. Rizon, R. Nagarajan, and S. Yaacob, "Inferring of human emotional states using multichannel EEG," *Europ. Journal of Scientific Research*, vol. 48, no. 2, pp. 281–299, 2010.

[147] R. Khosrowabadi, H. Quek, A. Wahab, and K. Ang, "EEG-based emotion recognition using self-organizing map for boundary detection," in *Int. Conf. on Pattern Recognition*, 2010, pp. 4242–5.

[148] A. Vijayan, D. Sen, and A. Sudheer, "EEG-based emotion recognition using statistical measures and auto-regressive modeling," in *IEEE Int. Conf. on Computational Intelligence & Communication Technology*, 2015, pp. 587–591.

[149] L. Schmidt and L. Trainor, "Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions," *Cognition and Emotion*, vol. 15, no. 4, pp. 487–500, 2001.

[150] M.-S. Park, H.-S. Oh, H. Jeong, and J.-H. Sohn, "EEG-based emotion recogntion during emotionally evocative films," in *Int. Winter Workshop on Brain-Computer Interface*, 2013, pp. 56–57.

[151] A. Choppin, "EEG-based human interface for disabled individuals: Emotion expression with Neural Networks," Master Thesis, Tokyo Institute of Technology, 2000.

[152] Z. Khalili and M. Moradi, "Emotion detection using brain and peripheral signals," in *IEEE Int. Biomedical Engineering Conf.*, 2008, pp. 1–4.

[153] S. Makeig, G. Leslie, T. Mullen, D. Sarma, N. Bigdely-Shamlo, and C. Kothe, "First demonstration of a musical emotion BCI," in *Int. Conf. on Affective Computing and Intelligent Interaction (ACII)*, 2011, pp. 487–496.

[154] F. Bahari and A. Janghorbani, "EEG-based emotion recognition using recurrence plot analysis and k nearest neighbor classifier," in *IEEE Iranian Conf. on Biomedical Engineering (ICBME)*, 2013, pp. 228–233.

[155] X. Jie, R. Cao, and L. Li, "Emotion recognition based on the sample entropy of EEG," *Bio-Medical Materials and Engineering*, vol. 24, no. 1, pp. 1185–1192, 2014.

[156] Y.-H. Liu, W.-T. Cheng, and Y.-T. Hsiao, "EEG-based emotion recognition based on kernel Fisher's discriminant analysis and spectral powers," in *IEEE Int. Conf. on Systems, Man and Cybernetics*, 2014, pp. 2221–5.

[157] M. Karg, K. Kühnlenz, and M. Buss, "Towards a system-theoretic model for transition of affect," in *AISB Convention - Mental States, Emotions and their Embodiment*, 2009, pp. 24–31.

[158] R. Levenson, "Emotion and the autonomic nervous system: A prospectus for research on autonomic specificity," in *Social Psychophysiology and Emotion: Theory and Clinical Applications*, H. Wagner, Ed., Wiley, 1988, pp. 17–42.

[159] I. Mauss and M. Robinson, "Measures of emotion: A review," *Cognition & Emotion*, vol. 23, no. 2, pp. 209–237, 2009.

[160] W.-H. Wen, G.-Y. Liu, N.-P. Cheng, J. Wei, P.-C. Shangguan, and W.-J. Huang, "Emotion recognition based on multi-variant correlation of physiological signals," *IEEE Trans. on Affective Computing*, vol. 5, no. 2, pp. 126–140, 2014.

[161] D. Sander, D. Grandjean, and K. Scherer, "A systems approach to appraisal mechanisms in emotion.," *Neural Networks*, vol. 18, no. 4, pp. 317–52, 2005.

[162] M. Karg, S. Haug, K. Kuhnlenz, and M. Buss, "A dynamic model and system-theoretic analysis of affect based on a Piecewise Linear system," in *IEEE Int. Symp. on Robot and Human Interactive Communication*, 2009, pp. 238–244.

[163] A. Hakim, S. Marsland, and H. Guesgen, "Computational analysis of emotion dynamics," in *IEEE Int. Conf. on Affective Computing and Intelligent Interaction*, 2013, pp. 185–190.

[164] E. Hudlicka, "Affective computing for game design," in *Int. N.A. Conf. on Intelligent Games and Simulation*, 2008, pp. 5–12.

[165] C. Waugh, P. Hamilton, and I. Gotlib, "The neural temporal dynamics of the intensity of emotional experience," *Neuroimage*, vol. 49, no. 2, pp. 1699–1707, 2010.

[166] S. Daselaar, H. Rice, D. Greenberg, R. Cabeza, K. LaBar, and D. Rubin, "The spatiotemporal dynamics of autobiographical memory: Neural correlates of recall, emotional intensity, and reliving," *Cerebral Cortex*, vol. 18, no. 1, pp. 217–29, 2008.

[167] C. Kohler, T. Turner, W. Bilker, C. Brensinger, S. Siegel, S. Kanes, R. Gur, and R. Gur, "Facial emotion recognition in schizophrenia: Intensity effects and error pattern," *The American Journal of Psychiatry*, vol. 160, no. 10, pp. 1768–74, 2003.

[168] V. Orgeta and L. Phillips, "Effects of age and emotional intensity on the recognition of facial emotion," *Experimental Aging Research*, vol. 34, no. 1, pp. 63–79, 2007.

[169] J. Bakker, M. Pechenizkiy, and N. Sidorova, "What's your current stress level? Detection of stress patterns from GSR sensor data," in *IEEE Int. Conf. on Data Mining Workshops*, 2011, pp. 573–580.

[170] G. Valenza, L. Citi, A. Lanata, E. P. Scilingo, and R. Barbieri, "A nonlinear heartbeat dynamics model approach for personalized emotion recognition," in *IEEE Int. Conf. of the Engineering in Medicine and Biology Society*, 2013, pp. 2579–82.

[171] D. Kulic and E. Croft, "Affective state estimation for human-robot interaction," *IEEE Trans. on Robotics*, vol. 23, no. 5, pp. 991–1000, 2007.

[172] J. Bailenson, E. Pontikakis, I. Mauss, J. Gross, M. Jabon, C. Hutcherson, C. Nass, and O. John, "Real-time classification of evoked emotions using facial feature tracking and physiological responses," *Int. Journal of Human-Computer Studies*, vol. 66, no. 5, pp. 303–317, 2008.

[173] S. Kim, F. Valente, M. Filippone, and A. Vinciarelli, "Predicting continuous conflict perception with bayesian Gaussian processes," *IEEE Trans. on Affective Computing*, vol. 5, no. 2, pp. 187–200, 2014.

[174] C. Roether, L. Omlor, A. Christensen, and M. Giese, "Critical features for the perception of emotion from gait," *Journal of Vision*, vol. 9, no. 6, pp. 1–32, 2009.

[175] M. Wöllmer, F. Eyben, S. Reiter, B. Schuller, C. Cox, E. Douglas-Cowie, and R. Cowie, "Abandoning emotion classes-towards continuous emotion recognition with modelling of long-range dependencies," in *INTERSPEECH*, 2008, pp. 597–600.

[176] M. Korhonen, D. Clausi, and E. Jernigan, "Modeling emotional content of music using system identification.," *IEEE Trans. on Systems, Man, and Cybernetics, Part B*, vol. 36, no. 3, pp. 588–99, 2006.

[177] G. Laurans, P. Desmet, and P. Hekkert, "The emotion slider: A self-report device for the continuous measurement of emotion," in *IEEE Int. Conf. on Affective Computing and Intelligent Interaction*, 2009, pp. 408–13.

[178] R. Cowie and E. Douglas-Cowie, "'FEELTRACE': An instrument for recording perceived emotion in real time," in *Speech and Emotion*, 2000, pp. 19–24.

[179] R. Cowie, M. Sawey, C. Doherty, J. Jaimovich, C. Fyans, and P. Stapleton, "Gtrace: General trace program compatible with EmotionML," in *IEEE Int. Conf. on Affective Computing and Intelligent Interaction*, 2013, pp. 709–710.

[180] F. Nagel, R. Kopiez, O. Grewe, and E. Altenmüller, "EMuJoy: Software for continuous measurement of perceived emotions in music," *Behavior Research Methods*, vol. 39, no. 2, pp. 283–290, 2007.

[181] A. Metallinou and S. Narayanan, "Annotation and processing of continuous emotional attributes: Challenges and opportunities," in *IEEE Int. Conf. and Workshops on Automatic Face and Gesture Recognition*, 2013, pp. 1–8.

[182] S. Kreibig, G. Schaefer, and T. Brosch, "Psychophysiological response patterning in emotion: Implications for affective computing," in *A Blueprint for an Affectively Competent Agent*, K. Scherer, T. Bänzinger, and E. Roesch, Eds., Oxford University Press, 2010, pp. 105–130.

[183] L. Ljung, *System identification*. Prentice Hall PTR, 1999.

[184] B. Meuleman and K. Scherer, "Nonlinear appraisal modeling: An application of machine learning to the study of emotion production," *IEEE Transactions on Affective Computing*, vol. 4, no. 4, pp. 398–411, 2014.

[185] G. Schöner, "Dynamical systems approaches to cognition," in *Cambridge Handbook of Computational Psychology*, R. Sun, Ed., Cambridge University Press, 2008, pp. 101–126.

[186] Y. Sandamirskaya, "Dynamic neural fields as a step toward cognitive neuromorphic architectures.," *Frontiers in Neuroscience*, vol. 7, no. 276, pp. 1–13, 2014.

[187] S. Amari, "Dynamics of pattern formation in lateral-inhibition type neural fields," *Biological Cybernetics*, vol. 27, pp. 77–87, 1977.

[188] C. Rasmussen, "Gaussian processes for machine learning.," *Machine Learning, LNAI*, vol. 3176, pp. 63–71, 2003.

[189] M. Stone, "An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion," *Journal of the Royal Statistical Society, Series B*, vol. 39, no. 1, pp. 44–47, 1977.

[190] S. Watanabe, "Asymptotic equivalence of Bayes cross-validation and widely applicable information criterion in singular learning theory," *Journal of Machine Learning Research*, vol. 11, pp. 3571–3594, 2010.

[191] Y.-H. Yang, Y.-C. Lin, Y.-F. Su, and H. Chen, "Music emotion classification: A regression approach," in *IEEE Int. Conf. on Multimedia and Expo*, 2007, pp. 208–211.

[192] M. Soleymani, G. Chanel, J. Kierkels, and T. Pun, "Affective characterization of movie scenes based on multimedia content analysis and user's physiological emotional responses," in *10th IEEE Int. Symposium on Multimedia*, 2008, pp. 228–235.

[193] H. Critchley, S. Wiens, P. Rotshtein, A. Ohman, and R. Dolan, "Neural systems supporting interoceptive awareness.," *Nature Neuroscience*, vol. 7, no. 2, pp. 189–95, 2004.

[194] *http://arduino.cc/en/Main/ArduinoBoardUno*. Arduino UNO, technical details, accessed March, 2015.

Supervised students' work

[195] M. Lodhi, "Multivariate feature selection: A comparison of search methods to find interactions of features," Master thesis, Technische Universität München, 2014.

[196] M. Debout, "Multivariate feature selection for machine learning," Advanced Seminar, Technische Universität München, 2014.

[197] F. Dörning, "Emotion recognition using Neural Networks: from discrete to continuous classification," Bachelor thesis, Technische Universität München, 2014.

[198] M. Myat Mon, "Entropy-based feature selection on affective physiological data," Bachelor thesis, Technische Universität München, 2013.

[199] H. Khelil, "Temporal analysis of emotional EEG and physiological data," Bachelor thesis, Technische Universität München, 2012.

Author's publications

[200] R. Jenke and A. Peer, "Towards incorporating appraisal into emotion recognition: A dynamic model for intensity estimation from physiological signals," *under review*, 2015.

[201] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG," *IEEE Trans. on Affective Computing*, vol. 5, no. 3, pp. 327–339, 2014.

[202] R. Jenke, A. Peer, and M. Buss, "Effect-size-based electrode and feature selection for emotion recognition from EEG," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013, pp. 1217–1221.

[203] R. Jenke, A. Peer, and M. Buss, "A comparison of evaluation measures for emotion recognition in dimensional space," in *IEEE Int. Conf. on Affective Computing and Intelligent Interaction*, 2013, pp. 822–826.

[204] N. Martens, R. Jenke, M. Abu-alqumsan, C. Kapeller, C. Hinterm, C. Guger, A. Peer, and M. Buss, "Towards Robotic Re-Embodiment using a Brain-and-Body-Computer Interface," in *IEEE Int. Conf. on Intelligent Robots and Systems (IROS)*, 2012, pp. 5131–5132.

[205] M. Abu-alqumsan, R. Jenke, A. Peer, and M. Buss, "Robotic Re-Embodiment using a Brain-Computer Interface," in *Neuroscience*, New Orleans, USA, 2012.

[206] M. Abu-Alqumsan, N. Martens, R. Jenke, C. Kapeller, C. Hintermüller, A. Peer, and M. Buss, "Adaptive Brain-Computer Interface and Its Application in Robotics," in *4th Workshop for Young Researchers on Human-Friendly Robotics, Twente, The Netherlands*, 2011.

[207] R. Jenke, "Analysis of the relation between Bayes classification and statistical testing," Diplomarbeit, Technische Universität München, 2010.

[208]   M. Karg, R. Jenke, W. Seiberl, K. Kühnlenz, A. Schwirtz, and M. Buss, "A comparison of PCA, KPCA and LDA for feature extraction to recognize affect in gait kinematics," in *IEEE Int. Conf. on Affective Computing and Intelligent Interaction*, 2009, pp. 1–6.