Modeling Affective State using Learning Vector Quantization

Gert-Jan de Vries

Modeling Affective State using Learning Vector Quantization

Gert-Jan de Vries

Cover design The cover design was created by composition of the learning dynamics of 3 by 2 by 3 RSLVQ prototypes on a simulated dataset consisting of 3 classes, where two classes (orange and green) are represented each by a single Gaussian cluster and a third class (blue) by a mixture of two smaller Gaussian clusters. Overlayed are mannequin-shaped heatmaps that indicate the location at which people feel the following emotions within their body (pairwise top,bottom): Happiness, Love (orange); Surprise, Envy (blue); Pride, Anger (green). The green representations were created by exchanging the red and green color channels. The six mannequins were taken from Nummenmaa et al. (2013) and printed with permission of Prof. Nummenmaa.

rijksuniversiteit
groningen

Modeling Affective State using Learning Vector Quantization

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. E. Sterken
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

vrijdag 28 november 2014 om 12.45 uur

door

Jan Johannes Gerardus de Vries
geboren op 12 maart 1984
te Stadskanaal

Promotor
Prof. dr. M. Biehl

Copromotor
Dr. S. C. Pauws

Beoordelingscommissie
Prof. dr. G. Bhanot
Prof. dr. T. Martinetz
Prof. dr. N. Petkov

ISBN: 978-90-367-7388-1

Inhoudsopgave

Acknowledgments

Many people contributed to the research leading to this thesis, whom I would like to thank. First of all, I would like to thank Michael Biehl for introducing me to the topic of LVQ by suggesting an analytical study of RSLVQ as topic for my Master's thesis. In hindsight I should, afterwards, have taken the opportunity you gave me to continue researching LVQ in a subsequent PhD at Groningen University in 2007, however I chose for another challenge of living on the other side of the country and working in industrial research at Philips Research. There I met Joyce Westerink and Martin Ouwerkerk, who very passionately introduced me to the field of emotions, physiology and psychology; all scary words for a mathematically oriented computer scientist. I took up the challenge and together with Joyce, Martin, Ramon Clout, Jack van den Eerenbeemd, Roos Rajae, Evelijne Hart De Ruijter, Peter Sels and Erik Bos measured and studied various affective states using unobtrusive measurements of physiological signals. I would like to thank them all for being a wonderful multidisciplinary team to work in. Joyce, thank you for providing, as project leader, the right balance of freedom and guidance to let me explore and grow as a researcher. You have further inspired me to perform research there where multiple disciplines meet. Martin, thank you for sharing your expertise in the hardware on which my software would run, for letting me use your lab and for your help when connectors failed and prototypes broke down. Ramon, thank you for the inspirational collaboration on creating various algorithms and demonstrator systems for interpreting physiological signals.

Over the years, various other projects on physiology, emotions, stress, vitality and sleep followed, where I collaborated with many inspiring people from a wide range of disciplines. I would like to thank these colleagues at Philips Research for providing a very pleasant working atmosphere and inspiring interactions during both project related collaborations, but also when simply having a cup of tea. In particular, I thank Paul Lemmens and Dirk Brokken for their help in understanding and applying appraisal theory. Caifeng Shan and Vincent Jeanne, thank you very much for letting me use your code for feature extraction from facial pictures and answering nitty gritty detail questions even years after you worked on the topic. Thanks to

Stijn de Waele for his work on implementing algorithms for the preprocessing and interpretation of ECG and Joris Janssen for his help in creating a complete overview of academic literature on emotion recognition from various modalities.

Somewhat regretting my decision not to pursue a PhD after finishing my study, the plan arose to bundle the research done in the form of a PhD. I am thankful that, when I contacted Michael in 2010, he was very open to the idea and agreed to be my promotor. Michael, thanks for your advise and guidance throughout the years, for always being available to brainstorm about analyses and the interpretation of results, and for the wonderful dinners you organized for Steffen and me in Groningen. That brings me to Steffen Pauws, my local supervisor. Thank you for your support throughout the project, for the many reviews performed, your tips to improve the various texts and for joining me on the trips to Groningen even when that meant we had to leave from Eindhoven station prior to 7 AM.

The context in which I performed my PhD research, was next to a full time job. To complicate matters further, in 2011 I moved within Philips Research to a completely new domain: Healthcare, more specifically readmission management for chronically ill patients. The job of analyzing clinical data posed yet another multidisciplinary challenge, which I enjoyed exploring with Aleksandra Tesanovic and Gijs Geleijnse. Combining all activities was far from easy and from time to time I seriously considered stopping my PhD. I want to thank Aleksandra, Gijs, Igor Berezhnoy and Marjolein van der Zwaag for their support in providing motivation to continue and maintain the right focus to bring this endeavour to a good end. Aleksandra, a very special thanks to you for your trust in me being able to handle the challenges in parallel, for letting me be part of your team and supporting me throughout. Igor, thanks for relentlessly 'bugging' me and for your unique and humorous support to not let me quit. Thank you, Jon Mason, for helping out with the cover design. Ioanna Sokoreli and Marjolein, thank you for being my paranymphs and for your support in the final stage of the project.

I would like to thank Ad Denissen, Marian Dekker and Geert van Boxtel for providing access to the dataset used for stress recognition from physiology in this thesis and for their help in interpreting the results. Anarta Ghosh and Aree Witoelar, thank you for helping me with the analysis of RSLVQ and in extending the work to other LVQ methods. I want to thank the members of the reading committee for their effort in reviewing my thesis: Prof. Gyan Bhanot, Prof. Thomas Martinetz, and Prof. Nicolai Petkov.

I am grateful to the Intelligent Systems group at Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen for hosting me as external PhD student and to Philips Research management, including all group leaders I had over the years, for their support, allowing the use of data and the technical infrastructure to perform analyses.

Also thanks to my room mates at Philips Research: Aleksandra, Gijs, Wim Stut, Paul ten Brink and Lukas Gorzelniak who made working days both productive and enjoyable while leaving the energy to work on my PhD in the evenings. Paul, a special thanks to you for helping me improve the Dutch translation of the thesis summary, for sharing my passion for classical music and for your magical humor, making even Monday mornings feel like Friday afternoon, "Stop it!". Thanks to all colleagues and friends whom I have not personally mentioned yet, but did contribute to my academic development.

Finally, a special thanks goes to my family, and my parents in particular, for supporting my career in all non-scientific ways. There was always a room available in 'Hotel De Vries' when I combined a visit to the university with a weekend at my parental home. Also thanks to my two-wheeled stallion for letting me experience many different emotions and bringing new ideas and inspiration during long-distance cycling. Last but not least I want to express my gratitude to the composers of the romantic era: Frédéric Chopin, Antonín Dvořák, Johannes Brahms and Sergei Rachmaninoff, for providing emotional inspiration through their wonderful music. In particular the piano concerto's by Chopin and Rachmaninoff supported writing of my thesis a lot.

Gert-Jan de Vries
Eindhoven
October 1, 2014

Abbreviations and Symbols

Nomenclature

$\xi_\sigma \in \mathbb{R}^d, y_\sigma$	d-dimensional input sample indexed by σ, representing class y_σ
y	label of input sample
$\mathbf{w}_S \in \mathbb{R}^d, c_S$	d-dimensional prototype indexed by S, representing class c_S
d_S	Euclidian distance between a sample ξ and prototype $\mathbf{w}_S$
N_c	number of classes
ϵ_g	generalization error

Acronyms

Psycho-physiology

ANS	Autonomous Nervous System
BP	Blood Pressure
BVP	Blood Volume Pulse
CNS	Central Nervous System
CPM	Component Process Model
ECG	Electrocardiogram
EDA	Electrodermal Activity
EEG	Electroencephalogram
EMG	Electromyogram
fMRI	functional Magnetic Resonance Imaging

GSR	Galvanic Skin Response
HF	High Frequency
HRV	Heart Rate Variability
IBI	Inter-Beat Interval
LF	Low Frequency
MRI	Magnetic Resonance Imaging
PAD	Pleasure, Arousal, Dominance
PNN50	Proportion of Inter-Beat Intervals (IBIs) > 50 ms
PNS	Peripheral Nervous System
PPG	Photoplethysmography
PSNS	Parasympathetic Nervous System
RMSSD	Root Mean Square of Successive Differences
RSA	Respiratory Sinus Arrhythmia
RSP	Respiration
SA	sinoatrial
SC	Skin Conductance
SCL	Skin Conductance Level
SCR	Skin Conductance Response
SDNN	Standard Deviation of IBIs
SDSD	Standard Deviation of Successive Differences
SNS	Sympathetic Nervous System
ST	Skin Temperature
VAD	Valence, Arousal, Dominance
VHF	Very High Frequency
VLF	Very Low Frequency

Machine learning

ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
ARMA	Autoregressive-Moving Average
AUC	Area Under the Curve
BLD	Best Linear Decision
BN	Bayesian Network
DT	Decision Tree
FDA	Fisher Discriminant Analysis
GLVQ	Generalized Learning Vector Quantization

GMLVQ	Generalized Matrix Learning Vector Quantization
GRLVQ	Generalized Relevance Learning Vector Quantization
HMM	Hidden Markov Model
KFDA	Kernel Fisher Discriminant Analysis
KLDA	Kernel Linear Discriminant Analysis
kNN	k-Nearest Neighbors
KPCA	Kernel Principal Component Analysis
LDA	Linear Discriminant Analysis
LFM	Learning From Mistakes
LFM-W	Learning From Mistakes with a window
LVQ	Learning Vector Quantization
MRSLVQ	Matrix Robust Soft Learning Vector Quantization
NBN	Naive Bayesian Network
NKFDA	Non-linear Kernel Fisher Discriminant Analysis
NLP	Natural Language Processing
ODE	Ordinary Differential Equations
PCA	Principal Component Analysis
PNN	Probabilistic Neural Network
QDC	Quadratic Discriminant Classifier
RBF	Radial Basis Function
RF	Random Forest
ROC	Receiver Operating Characteristic
RSLVQ	Robust Soft Learning Vector Quantization
RT	Regression Tree
SVC	Support Vector Classifier
SVM	Support Vector Machine
VQ	Vector Quantization

Image processing

AU	Action Unit
FACS	Facial Action Coding System
HLAC	Higher-order Local Autocorrelation
HLACLF	HLAC-like features
LBP	Local Binary Patterns
LDP	Local Directional Patterns
SIFT	Scale-Invariant Feature Transform

Hoofdstuk 1

GENERAL INTRODUCTION

Each research field has interesting challenges, but most fascinating challenges arise where multiple research fields meet. This thesis cross-fertilizes various research fields from life-sciences, including psychology and physiology, with computing science and artificial intelligence. By studying and embedding machine learning techniques in affective sciences, i.e., those that study human emotion, this thesis outlines advances in the recognition of human affect using various measurement modalities.

From a technical perspective, we employ and study learning dynamics of machine learning techniques. These methods can be subdivided into supervised and unsupervised learning. The former type aims at separating data from different classes by learning from labelled samples, the latter aims at finding optimal separation of unlabelled data samples into clusters that represent the overall data structure well. Unsupervised learning is beyond the scope of this thesis. Classification methods can be further subdivided into black-box and white-box methods. Black-box methods can be observed through input and output, but do not have an insightful internal structure. White-box methods, on the contrary, allow for more natural inspection of their internals and thereby provide insights into what data properties separate each class from the others. In this thesis, we primarily focus on Learning Vector Quantization (LVQ) which is a typical white-box method. Amongst the black-box methods are many of the popular classification techniques such as Support Vector Machine (SVM), Artificial Neural Network (ANN), k-Nearest Neighbors (kNN) as well as various ensemble methods such as Random Forest (RF). There have been developed techniques that can extract information from such classifiers (Tickle et al. 1998, Martens et al. 2007, Cortez and Embrechts 2013), however they do require significant additional analysis. For that reason white-box methods could be preferable over black-box methods, hence our interest to study the learning dynamics of LVQ. We will include several black-box methods in our experiments for comparison, given their popularity in other studies.

From the perspective of the application domain, we study the recognition of

various affective states from a variety of measurable quantities. The research on affect has a long history in psychology. The book by Charles Darwin titled *On the Expression of the Emotions in Man and Animals* is considered the pioneering work (Darwin 1872). Since then the topic has been studied widely in controlled (laboratory) conditions. With the recent technological developments that enabled miniaturization of measurement equipment also studies of emotion in daily life became feasible (Picard and Scheirer 1999, Westerink et al. 2009). Traditionally, psychological theories were developed and tested in experiments. In recent years the field of emotions was also discovered by computing science as an area with challenging problems and data driven studies, such as studied in this thesis became possible. The term *Affective Computing* for this multidisciplinary field was first introduced by Rosalind Picard in 1995 (Picard 1995), but the notion of monitoring one's emotional state with the help of computers originates in the 1970's (Kuechenmeister et al. 1970).

1.1 Scope of this study

Whereas a substantial part of affective research focusses on the fundamental research of developing theories on underlying mechanisms involved in emotion expression or detection, this study focusses on applying such knowledge in systems for affect recognition. Although this work is not the first to address affect recognition, it does comprise pioneering work in the application of white-box methods, in particular the prototype based LVQ methods to affective computing. The application of such methods allows for a deeper study of the aspects that most influence the automated affect recognition process and could potentially lead to a better understanding of the processes involved in human affect. In order to reflect the diversity in approaches to study affective phenomena, three of the following chapters will approach affective computing from different angles. In addition, we perform a theoretical study of various LVQ methods to better understand the learning dynamics involved.

1.2 Outline

In the following chapter we will further define and describe the technical and domain concepts that are studied in this thesis. After that, Chapter 3 will present a theoretical study of the performance and other properties of various LVQ variants

during the training phase of these classifiers. The three chapters following that will describe different affective classifiers using a variety of input modalities, thereby approaching affective computing from three different perspectives: Stress detection from physiological measurements reflecting the bodily perspective (Chapter 4), facial expression recognition using facial images reflecting the facial perspective (Chapter 5), and emotion detection using mental appraisals reflecting the cognitive perspective (Chapter 6). Chapter 7 will describe potential applications that make use of the various affective classifiers. Finally, a summary will be given in Chapter 8.

Hoofdstuk 2

PROBLEM DEFINITIONS AND METHODS

Measuring affect requires an understanding of the affective domain as well as the technology needed to do so. This chapter provides an introduction into the concepts used and studied in this thesis, both from the technical as well as the application domain perspective. Section 2.1 introduces the concept of affect in general, Section 2.2 describes the modalities used to measure various aspects of affect and how the signals obtained through these measurements are preprocessed to form feature sets that are used and studied in the following chapters. Further background, societal relevance and details of features used for each of the affective sub-domains will be provided in their respective chapters: Stress detection from physiological measurements (Chapter 4), facial expression recognition using facial images (Chapter 5), and emotion detection using mental appraisals (Chapter 6). Section 2.3 introduces the classification techniques studied and used in this thesis.

2.1 Definition of Affect

Affect is generally considered a ”superordinate label for all ”emotional” phenomena” (Kleinginna Jr and Kleinginna 1981), which covers concepts as Emotion, Mood, Attitudes and Personality. These different affective phenomena are ordered by decreasing intensity and event focus, and increasing duration (Scherer 2005). Defining *Emotion* has been topic of debate for over 130 years already, up to the point that Young concluded that ”almost everyone except the psychologist knows what an emotion is...” (Young 1973). Depending on the point of view used to study emotions, different definitions are used. A comprehensive overview of definitions was provided in 1981 consisting of no less than 92 definitions (Kleinginna Jr and Kleinginna 1981). Since the first definition, given by William James in 1884, ”the bodily changes follow directly the PERCEPTION of the exciting fact, and that our feeling of the same changes as they occur IS the emotion” (James 1884), many different definitions have been posed, but have still not lead to consensus on a single definition of emotion (Scherer 2005). On a high level, there can be distinguished two types of definitions: top-down and bottom-up generation of emotion (McRae et al. 2012).

The bottom-up view is in line with the original definition of James and argues that an initial subconscious reaction to stimuli coming from the environment causes bodily reactions and only then we become consciously aware of the emotion. On the other hand, the top-down approach reasons that we first create a mental awareness of emotions only after which the bodily reactions are triggered. More recently both views have been integrated into a view that consider the top-down and bottom-up processes as parallel (Scherer et al. 2001), for which also neurological evidence has been found (Ochsner et al. 2009). Chapters 4 and 5 follow the bottom-up view as they aim at the detection of affective states from measurements of bodily and facial reactions, while Chapter 6 approaches emotions through a top-down view using the cognitive processes as starting point.

Emotions are generally considered as short lasting (seconds to minutes) affective phenomena that have a clear identifiable cause and prepare the body for action, originating from the fight or flight reaction. To support sudden action taking, emotions lead to a variety of bodily responses such as raising heart rate and peripheral vasoconstriction to support the blood supply to muscles, as well as increased secretion of sweat on palms and soles to improve grip. Throughout human evolution more social emotions have developed, leading to a diverse palette of emotions. The bodily reactions enable the measurement of emotions without requiring self-report. In the following chapters, feasibility of automated detection of emotions from different measurement modalities will be explored.

Stress is a related concept to emotions, moods, or affect in general. Jones and Bright (2001) discuss that the layman's perspective of 'stress' confuses several concepts such as 'strain', 'pressure', 'demand' and 'stressors'. On the other hand, academic definitions are also quite diverse, although Cox (1993) disagrees with that assumption, he proposes the following broad definition: "Stress can be defined as a psychological state which is part of and reflects a wider process of interaction between individuals and their work environment" (Cox 1993). He also distinguishes three approaches to stress: the engineering (or contextual), physiological, and psychological approach. In such higher level view stress can be seen as an affective phenomenon as it comprises psychological and physiological reflections of external stressors.

While some definitions of Affect are centered around the bodily reactions to themselves, the general view is that Affect is defined in terms of the personal evaluation or experience (either fully in terms of cognitive aspects, or in terms of the cognitive awareness of the bodily reactions). Ground truth on emotion is therefore

most often gathered by self-report, measured through questionnaires. A disadvantage of this means of ground truth is the subjectiveness that originates from different interpretations of the questions, different frames of reference and effects like giving socially desired answers, all of which add noise to the ground truth labeling. Ground truth can also be defined in terms of the stimuli presented (e.g., a labeling the emotional response to a fearful stimulus with 'fear'). While this technique is objective, it fully takes away the personal differences in emotional interpretation of the stimuli. In sum there is no generally preferable means to gather emotional ground truth, hence we chose the most appropriate method for each of the experiments outlined in the following chapters.

2.2 Measuring Affect

The human mind and body are connected through the nervous system, which consists of a central and peripheral part. The Central Nervous System (CNS), formed by the brain and spinal cord, can be seen as a central coordinator which receives and processes input from the body and outputs instructions. The Peripheral Nervous System (PNS) serves as two-way communication means between CNS and various body parts, all the way to the extremities. Within these two closely collaborative nervous systems, there are various subsystems. Many bodily responses, among which affective responses, are regulated by the Autonomous Nervous System (ANS). The ANS steers physiological processes, as illustrated in Figure 2.1, by two main subsystems: the Sympathetic Nervous System (SNS) and Parasympathetic Nervous System (PSNS). The SNS can be seen as a "quick response mobilising system" and the parasympathetic as a "more slowly activated dampening system" (McEwen et al. 2001). The PSNS is responsible for maintaining the basic organ functions in a rest state by regulating physiological functions such as blood-flow, digestion and excretion of various fluids (e.g., tears, sweat). The SNS, on the other hand, enables fast activation of the *fight-or-flight response* (Cannon 1967). A fine balance between these counteracting systems manages a stable bodily state termed homeostasis.

Although many of the processes regulated by the ANS affect the internal organs, there are several physiological processes of which resulting effects are externally observable through various measurement modalities. Among these are the activity of the heart, respiration system, sweat glands, temperature, and muscles. Next to the autonomous responses from the ANS, there are also regulatory processes of affect

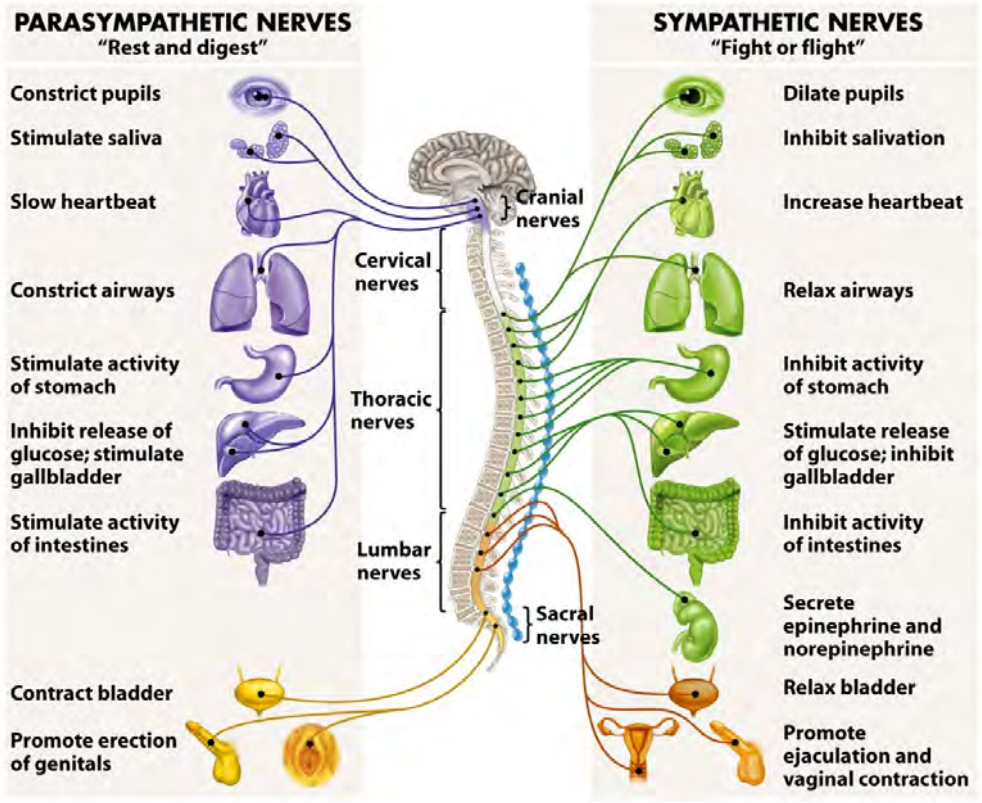

Figure 2.1: *Division of the ANS into PSNS and SNS (Freeman 2005) ©2005. Reprinted by permission of Pearson Education, Inc., Upper Saddle River, NJ.*

that involve awareness. Reflections of this can be seen in our behaviour, movements and (facial) expressions. In the following, several physiological modalities will be discussed, together with the measurement modalities that can be used to measure their effects. Chapters 4 to 6 will use these measurement modalities to gather measurements of various affective states.

2.2.1 Cardiac activity

By regulating the pace of cardiac activity, the flow of (oxygen rich) blood to the extremities can be regulated. Thereby, the body can meet the oxygen requirements of activated body parts. The heart consists of muscle tissue that contracts whenever

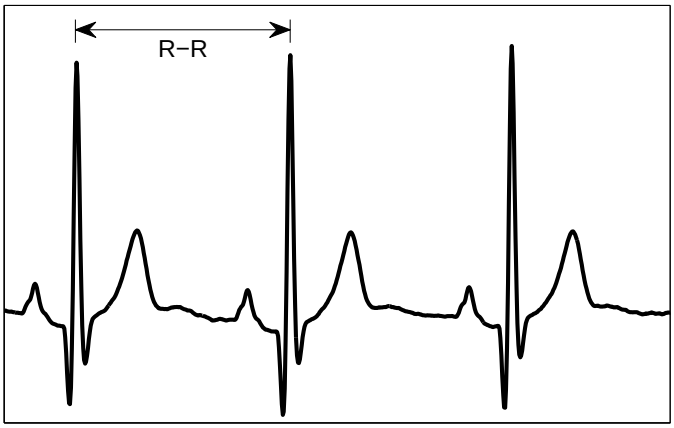

Figure 2.2: *Example trace of ECG recording representing three heart beats.*

activated through an electrical pulse. Starting in the sinoatrial (SA) node, an electrical pulse travels through the entire heart causing the four chambers to rhythmically contract. This electrical activity can be observed through Electrocardiogram (ECG) recordings that measure the fluctuations in electrical potential over at least two electrodes positioned on the skin and across the heart. Figure 2.2 shows an example trace of ECG. The highest peaks correspond to the contraction of the heart's ventricles and are termed R waves. The interval between successive R waves is called an R-R interval, or Inter-Beat Interval (IBI). Variations in IBI, also referred to as Heart Rate Variability (HRV) are known to reflect activity of SNS and PSNS, and can be analyzed in the time domain as well as in the frequency domain. The HRV guidelines paper (Malik et al. 1996) describes how different frequency intervals should be interpreted. Low frequency variations, in the interval 0.04 to 0.15 Hz, are known to vary with PSNS activity (Grossman and Taylor 2007), while high frequency variations (0.15-0.4 Hz) relate to SNS activity.

Preprocessing of the ECG signal consists of the following steps (as outlined in de Waele et al. (2009)): R-peak detection, IBI outlier removal, and HRV analysis. R-peak detection was performed using a pattern matching technique (Poor 1994). The resulting IBIs are filtered for outliers by using a sliding window histogram. In order to estimate frequency domain HRV features, an Autoregressive-Moving Average (ARMA) time series model is used to derive power in the frequency bands. Another unobtrusive means of measuring cardiac activity is Blood Volume Pulse (BVP), which measures pulses in the blood flow at a location characterized by good circulation, such as a finger or earlobe. Each heart beat creates such a pulse of blood which

can be measured optically with a Photoplethysmography (PPG) sensor. Compared to ECG, the peaks in the resulting signal are less sharp and the precision at which the IBIs can be measured is lower, in particular affecting measures of high frequency modulation of the heart rate.

2.2.2 Galvanic skin response

The main purpose of sweating is to cool down the body through evaporation. There is, however, also an emotional component that drives the sweat glands. Originating from the *fight-or-flight response* slightly moist palms and soles provided an advantage in reacting to dangers. This mechanism of ANS driven activation of sweat glands in reaction to various stressors can actually be observed all over the body but is strongest on the hands and feet (van Dooren et al. 2012) and is termed Galvanic Skin Response (GSR), Electrodermal Activity (EDA) or Skin Conductance (SC). The sweat glands consist of small ducts that can fill to various levels with fluid. Activation of the glands increases the moisture in the top level of the skin which results in a lower electrical resistance. The resistance can be measured by applying a small constant current and observing the (varying) voltage.

The resulting signal consists of a slow changing, tonic, component termed Skin Conductance Level (SCL) and a faster changing, phasic, response termed Skin Conductance Response (SCR) (Boucsein 2012). Figure 2.3 shows an example of such a GSR signal. In order to separate both components, the SCRs can be detected using a dedicated detection algorithm (Kohlish 1992). The tonic signal can be estimated by explicitly compensating for detected SCRs (de Vries and van der Zwaag 2010), but is often estimated from the raw SC signal. Features derived from GSR include levels and changes of the SCL and frequency, amplitude, rise time and half recovery time of SCR.

2.2.3 Respiratory activity

Changes in oxygen requirements of the body are regulated through the pace of breathing. In contrast to cardiac activity and GSR breathing is not completely autonomously regulated and can be partly influenced consciously on top of the regulation by the ANS. When the lungs fill with air, the thorax expands, which can be measured using a flexible chest strap containing stretchable conductive material of which the resistance increases the more it stretches. By measuring the voltage for an applied constant current this resistance can be derived.

Preprocessing starts with low-pass filtering (cut-off 0.5Hz), followed by an ana-

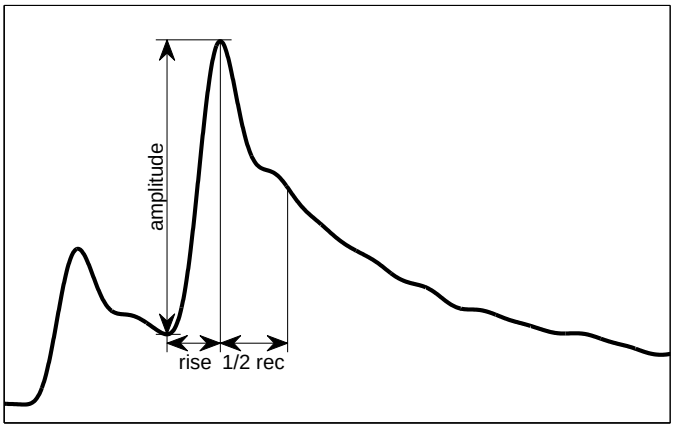

Figure 2.3: *Example trace of GSR recording representing two SCRs.*

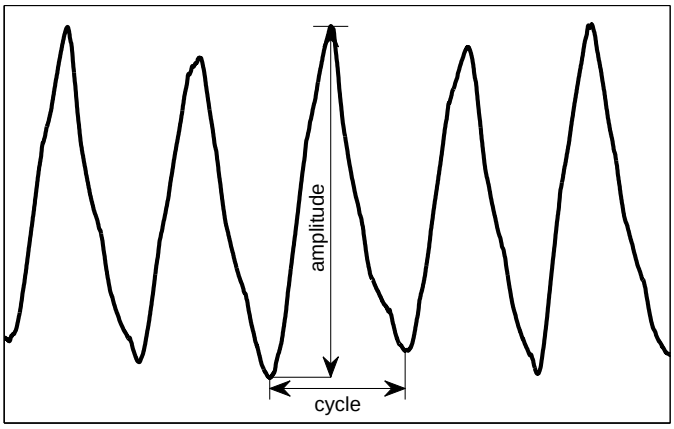

Figure 2.4: *Example trace of RSP recording representing five breaths.*

lysis for individual breaths. Using a localized min/max filter (Lemire 2006), local minima and maxima are detected. When found in the right order, they characterize a single breath. Based upon the distribution of identified breath amplitudes in a signal, too small or too large breaths (outliers) are removed. After this preprocessing the Respiration (RSP) signal is characterized by a sequence of breaths similar to the IBI signal for ECG, from which paces and amplitudes can be determined. Figure 2.4 shows an example trace of RSP, where a rising line indicates inhalation and a drop represents exhalation. Hence the signal between two valleys represents one breathing cycle.

2.2.4 Facial expressions

With the development of human kind to a more social creature, displaying emotions to others became more and more important. Facial expressions fulfill this need by activating certain facial muscles, which can be measured directly through facial Electromyogram (EMG) which measures the electrical activity of muscles between two electrodes. This, however, does require the attachment of various electrodes to the face which is far from comfortable and not socially accepted in daily life circumstances. As a solution, video of photo cameras, are used to record the resulting facial expressions. Using various image/video analysis techniques features can be extracted. In this thesis we used techniques that detect local textures such as edges and thereby are able to detect (de)formations of the face caused by facial expressions.

2.2.5 Cognitive processes

Whereas the above affect measures are mainly steered directly by the ANS, the CNS also plays a role in emotions. Although techniques such as Electroencephalogram (EEG) and functional Magnetic Resonance Imaging (fMRI) can be used to measure activity of certain parts of the brain, it is currently not possible to objectively measure emotional cognitive processes. To get insight into these processes, the golden standard of measurement are self-reported questionnaires. There are several cons to this measurement including the time needed to collect this information and the delay introduced between the cognitive process taking place and filling out the questionnaire, as well as subjective interpretation of the questions. The imprecision introduced needs to be taken into account when using this as input to an affective system, nevertheless these measurements provide an orthogonal view on the other measurements described above.

2.3 Classifiers

The work described in the following chapters will study and make use of several classification techniques. Main focus will be on Learning Vector Quantization (LVQ) methods; in addition several well known techniques of different nature will be used for comparison. In the following sections the classification methods used will be shortly introduced. In its simplest form, the classifiers can be described as a mapping of samples $\xi \in \mathbb{R}^N$ to classes $\widetilde{y}$. To that end, a training set of P_{tr} samples $\mathbf{\Xi}_{\text{tr}} = [\xi_1, \xi_2, \ldots, \xi_{P_{\text{tr}}}]^\top$ and their respective labels $\mathbf{Y}_{\text{tr}} = [y_1, y_2, \ldots, y_{P_{\text{tr}}}]$ are available to infer knowledge from and a test set of P_{te} (unseen) samples $\mathbf{\Xi}_{\text{te}}$ is available to evaluate how well the predicted labels $\widetilde{\mathbf{Y}}_{\text{te}} = [\widetilde{y}_1, \widetilde{y}_2, \ldots, \widetilde{y}_{P_{\text{te}}}]$ represent the actual

labels $\mathbf{Y}_{\text{te}}$. The task of the classifiers is to optimize this mapping.

2.3.1 Learning Vector Quantization

Learning Vector Quantization (LVQ) comprises a family of classifiers that is of open box nature, that is, they provide direct insight into the information learned by the classifier. LVQ, initially proposed by Kohonen (1990), defines prototypes in the same (mathematical) space as the data to represent the classes. These prototypes are directly interpretable as they show characteristics of classes in terms of the features chosen to represents the input data. During training, samples are presented sequentially, and for each sample the closest prototype(s) are updated according to the following general framework in Equation (2.1). We will use the notation $\mathbf{w} \in \mathbb{R}^N$ for a prototype with its class labels c. J, K, S, T are used to index prototypes and their class labels; hence $(\mathbf{w}_J, c_J)$ represents a prototype with its corresponding class label.

In the on-line algorithm, examples are presented sequentially to the system and the prototypes are adapted by the following update step:

$$\mathbf{w}_S^{\mu} = \mathbf{w}_S^{\mu-1} + \frac{\eta}{N} f_S \left[d_1^{\mu}, \ldots, d_K^{\mu}, y_{\sigma}^{\mu}, \ldots\right] \left(\xi^{\mu} - \mathbf{w}_S^{\mu-1}\right), \tag{2.1}$$

where $\mathbf{w}_S^{\mu}$ denotes the prototype representing class S after presentation of μ examples and the learning rate η is rescaled with N. We use the shorthand f_S for the modulation function which controls, along with the learning rate η, the magnitude of the update of $\mathbf{w}_S$ towards or away from the current example. Typically, squared Euclidean distance is used as distance measure (d_T^{μ}), as defined in Equation (2.2).

$$d_T^{\mu} = (\xi^{\mu} - \mathbf{w}_T^{\mu})^{\top} (\xi^{\mu} - \mathbf{w}_T^{\mu}) \tag{2.2}$$

The number of prototypes can be chosen arbitrarily, at least one prototype per class. We have chosen to initialize the prototypes close to the center of the data, i.e., close to the position of all samples in the training set, according to:

$$\mathbf{w}_T^0 = \frac{1}{N} \sum_{i=1}^{N} \xi^i + \delta_T, \text{ with } \delta_T \sim N(0, \sigma), \tag{2.3}$$

where N is the amount of samples in the training set and $\sigma = 0.1$ was chosen to represent only a small deviation from the data center. In case of two-class classification, as e.g. in Chapter 4, an Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) can be obtained by considering various configurations by moving the prototypes $\mathbf{w}_S$ and $\mathbf{w}_T$ along the line connecting them. By doing so, the decision boundary is moved along this line, while remaining perpendicular.

Within this framework, different variants have been proposed, using different strategies for choosing the update strength and distance measure. We considered the following selection, which we implemented using a combination of Matlab (R2013a) and C++: LVQ 2.1, LFM, GLVQ, RSLVQ, GRLVQ and GMLVQ. The following sections will discuss these variants in further detail.

LVQ 2.1

LVQ 2.1 was proposed by Kohonen aiming at efficient separation between prototypes of different classes and has been shown to provide good classification results (Kohonen 1990, Neural Networks Research Centre, Helsinki 2002). Given an example ξ^μ, two nearest prototypes $\mathbf{w}_S$ and $\mathbf{w}_T$ are updated if the following conditions are met: (i) the classes c_S and c_T are different, and (ii) either c_S or c_T is equal to y^μ_σ. The prototype with the correct class is moved towards the data while the other is moved farther away with $f_S = 1, f_T = -1$ if $c_S = y^\mu_\sigma$; $f_S = -1, f_T = +1$ else.

It is well known that such learning rule has stability problems for unbalanced data sets, resulting in diverging prototypes with deteriorating performance (Kohonen 1990). Therefore, LVQ 2.1 restricts updates to examples ξ^μ which fall into a *window* around the decision boundary.

$$\min\left(\frac{d^\mu_T}{d^\mu_S}, \frac{d^\mu_S}{d^\mu_T}\right) > \rho, \quad \text{with} \quad \rho = \frac{1-\omega}{1+\omega} \tag{2.4}$$

where ω is a window parameter, $0 < \omega \leq 1$ and therefore $1 > \rho \geq 0$.

LFM

A simple modification to overcome the stability problems of LVQ 2.1 is restricting updates only on misclassified examples. Analogous to perceptron learning, we term this update rule as Learning From Mistakes (LFM). Here, the closest prototype $\mathbf{w}_J$ with the same class $c_J = y^\mu_\sigma$ (*correct winner*) and closest prototype $\mathbf{w}_K$ with a different class $c_K \neq y^\mu_\sigma$ (*incorrect winner*) are updated with $f_J = +1$ and $f_K = -1$, if the example is misclassified. On the contrary, if the winning prototype is already correct, the configuration is left unchanged. This prescription can be interpreted as a limiting case of the cost function based RSLVQ, which will be explained later in this section. Because the cost function of RSLVQ is bounded from below, stability can also be expected in LFM.

Generalized LVQ

Earlier LVQ prescriptions, including LVQ 2.1, were based on heuristic grounds. In contrast, a popular variant termed the Generalized Learning Vector Quantization

(GLVQ) was proposed in Sato and Yamada (1995) which introduced the cost function

$$E = \sum_{\mu} \Phi(\tau(\xi^{\mu})) \quad \text{with} \quad \tau(\xi^{\mu}) = C \cdot \frac{d_J^{\mu} - d_K^{\mu}}{d_J^{\mu} + d_K^{\mu}} \tag{2.5}$$

where $\Phi(\tau)$ is a (usually non-linear) monotonically increasing function, d_J is distance from the nearest correct prototype and d_K from the nearest incorrect prototype to the example ξ^{μ}. We insert the scaling parameter C which will be required for high dimensions. Stochastic gradient procedure on (2.5) yields the learning rule

$$f_J = 2C \frac{\partial \Phi(\tau)}{\partial \tau} \frac{d_K^{\mu}}{(d_J^{\mu} + d_K^{\mu})^2} \ , \quad f_K = -2C \frac{\partial \Phi(\tau)}{\partial \tau} \frac{d_J^{\mu}}{(d_J^{\mu} + d_K^{\mu})^2} \ . \tag{2.6}$$

where d_J is the distance to the closest correct prototype with matching class label and d_K the distance to the closest prototype with mismatching label. $\Phi(\tau)$ is a monotonically increasing function, often chosen to be a sigmoid function. We chose $\Phi(\tau)$ to be the sigmoid function described by Equation (2.7). This leads to the explicit update function in Equation (2.8).

$$\Phi(\tau) = \frac{1}{1 + exp(-\tau)} \tag{2.7}$$

$$\begin{aligned} \mathbf{w}_J^{\mu+1} &= \mathbf{w}_J^{\mu} + \eta \frac{\Phi'(\tau) d_K^{\mu}}{(d_J^{\mu} + d_K^{\mu})^2} (\xi^{\mu} - \mathbf{w}_J^{\mu}), \\ \mathbf{w}_K^{\mu+1} &= \mathbf{w}_K^{\mu} - \eta \frac{\Phi'(\tau) d_J^{\mu}}{(d_J^{\mu} + d_K^{\mu})^2} (\xi^{\mu} - \mathbf{w}_K^{\mu}), \end{aligned} \tag{2.8}$$

where the derivative of $\Phi(\tau)$ (Equation (2.7)) is given by:

$$\Phi'(\tau) = \Phi(\tau) * (1 - \Phi(\tau)) \tag{2.9}$$

Robust Soft LVQ

The Robust Soft Learning Vector Quantization (RSLVQ) algorithm (Seo and Obermayer 2003) was derived using a statistical modeling of the data and designed to overcome the stability problem of LVQ 2.1. RSLVQ introduces soft prototype assignments which act similarly to a soft window around the decision boundary. This algorithm minimizes a bounded cost function $E = -\ln(L)$ where L is based on a likelihood ratio function of a mixture model, described as

$$L = \prod_{\mu} \frac{p(\xi^{\mu}, \sigma^{\mu} | W)}{p(\xi^{\mu} | W)} \quad \text{with} \quad p(\xi^{\mu} | W) = \sum_{\sigma=1}^{N_{\sigma}} \sum_{j : c_j = \sigma} P_j \, p(\xi^{\mu} | j), \tag{2.10}$$

where $p(\xi^\mu|W)$ approximates the actual probability density $P(\xi)$, c.f. (3.1). It is assumed that every component j of the mixture generates examples which belong to one class, viz. c_j. N_σ is the number of classes and P_j is the probability that the examples are generated by a particular component j and $p(\xi^\mu|j)$ is the conditional probability that j generates a particular example ξ^μ.

The learning rule is obtained by performing stochastic gradient descent on the cost function E with respect to $\mathbf{w}_S$. We examine it for a Gaussian mixture ansatz as in Seo and Obermayer (2003), where it is chosen $p(\xi^\mu|j) = (2\pi v_j)^{(N/2)} \exp(-d_j^\mu/2v_j)$. Furthermore, every component is assumed to have equal probability $P(j) = 1/K$, $\forall j$ and equal variance $v_j = v_{\text{soft}}$, $\forall j$ where v_{soft} is called the softness hyperparameter. This gives the following modulation function

$$f_S = \frac{1}{v_{\text{soft}}} \begin{cases} P_\sigma(S|\xi^\mu) - P(S|\xi^\mu), & \text{if } c_S = y_\sigma^\mu \\ -P(S|\xi^\mu), & \text{else.} \end{cases} \tag{2.11}$$

with the assignment probabilities

$$P_\sigma(S|\xi^\mu) = \frac{\exp\left(-d_S^\mu/2v_{\text{soft}}\right)}{\sum_{j:c_j=\sigma^\mu} \exp\left(-d_j^\mu/2v_{\text{soft}}\right)}, \quad P(S|\xi^\mu) = \frac{\exp\left(-d_S^\mu/2v_{\text{soft}}\right)}{\sum_j \exp\left(-d_j^\mu/2v_{\text{soft}}\right)}, \tag{2.12}$$

see Seo and Obermayer (2003) for the derivations. $P_\sigma(S|\xi^\mu)$ describes the posterior probability that ξ^μ is assigned to the component S of the mixture, given that the example is generated by the correct class. $P(S|\xi^\mu)$ describes the posterior probability that ξ^μ is assigned to the component S of the complete mixture using all classes.

This yields the following explicit update step for RSLVQ:

$$\begin{aligned} \mathbf{w}_J^{\mu+1} &= \mathbf{w}_J^\mu + \frac{\eta}{v_{\text{soft}}} \left(P_J(S|\xi^\mu) - P(S|\xi^\mu)\right) (\xi^\mu - \mathbf{w}_J^\mu), \\ \mathbf{w}_K^{\mu+1} &= \mathbf{w}_K^\mu - \eta P(S|\xi^\mu)(\xi^\mu - \mathbf{w}_K^\mu), \end{aligned} \tag{2.13}$$

In experiments, we found that the softness in RSLVQ allows the prototypes to move away from cluster centers in the data. As long as the relative distances between prototypes and the decision boundary does not change, this will not affect the assignments made by the classifier. It, however, does allow the classifier to implicitly add more weight to certain dimensions by moving the prototypes away from the decision boundary in those dimensions. By doing so, it increases the contribution of that dimension to the calculation of distances between prototypes and samples as described in Equation (2.2). This observation allows the interpretation of the difference vector of pairs of prototypes (i.e., $\mathbf{w}_S - \mathbf{w}_T$) as an implicitly trained weight vector similar to the explicitly trained relevance vector of GRLVQ as will be described in the following section.

Generalized Relevance LVQ

As an extension to GLVQ, Generalized Relevance Learning Vector Quantization (GRLVQ) was proposed by Hammer and Villmann (2002), which assigns and trains a relevance per data dimension by integrating a relevance vector into the distance measure, i.e., Equation (2.2) is replaced by the following weighted distance measure:

$$d^{\mu}_{\lambda,T} = \sum_{i=1}^{N} \lambda^{\mu}_{i}(\xi^{\mu}_{i} - \mathbf{w}^{\mu}_{T,i})^2 \tag{2.14}$$

Both the prototypes and the relevance matrix are updated at the same time during training using an update scheme similar to GLVQ:

$$\begin{aligned} \mathbf{w}^{\mu+1}_{J} &= \mathbf{w}^{\mu}_{J} + 2\eta \frac{\Phi'(\tau_\lambda) d^{\mu}_{\lambda,K}}{(d^{\mu}_{\lambda,J} + d^{\mu}_{\lambda,K})^2} \lambda^{\mu}(\xi^{\mu} - \mathbf{w}^{\mu}_{J}), \\ \mathbf{w}^{\mu+1}_{K} &= \mathbf{w}^{\mu}_{K} - 2\eta \frac{\Phi'(\tau_\lambda) d^{\mu}_{\lambda,J}}{(d^{\mu}_{\lambda,J} + d^{\mu}_{\lambda,K})^2} \lambda^{\mu}(\xi^{\mu} - \mathbf{w}^{\mu}_{K}), \end{aligned} \tag{2.15}$$

where τ_λ, is defined as in Equation (2.5), replacing d^{μ}_{T} by $d^{\mu}_{\lambda,T}$. The update of the relevance vector is given by Equation (2.16), where $\mathbf{w}_S$ is the closest prototype.

$$\begin{aligned} \lambda^{\mu+1} = \lambda^{\mu} - 2\epsilon\Phi'(\tau_\lambda) \Bigg(& \frac{d^{\mu}_{\lambda,K} \cdot \lambda^{\mu-1}(\xi^{\mu} - \mathbf{w}^{\mu}_{J})}{(d^{\mu}_{\lambda,J} + d^{\mu}_{\lambda,K})^2} \\ & - \frac{d^{\mu}_{\lambda,J} \cdot \lambda^{\mu-1}(\xi^{\mu} - \mathbf{w}^{\mu}_{K})}{(d^{\mu}_{\lambda,J} + d^{\mu}_{\lambda,K})^2} \Bigg) \end{aligned} \tag{2.16}$$

The relevance vector λ was initialized using random samples drawn from on the interval $[0, 1]$, i.e., $\lambda^0 \sim U_N(0, 1)$, and normalized such that the elements of λ^0 sum up to 1.

Generalized Matrix LVQ

Generalized Matrix Learning Vector Quantization (GMLVQ), proposed by Schneider et al. (2009a), uses a generalized distance measure that takes into account relevances of individual, but also cross-relevances of multiple, data dimensions by extending the Euclidean distance measure with a $N \times N$ relevance matrix Λ, see Equation (2.17). Regularization reduces the learning capacity and is obtained by limiting the size (M) of the $M \times N$ matrix Ω. We choose $M = 2$ in our analyses.

$$d^{\mu}_{\Lambda,T} = (\xi^{\mu} - \mathbf{w}^{\mu}_{T})^{\top}\Lambda^{\mu}(\xi^{\mu} - \mathbf{w}^{\mu}_{T}), \quad \text{with } \Lambda^{\mu} = (\Omega^{\mu})^{\top}\Omega^{\mu} \tag{2.17}$$

Both the prototypes and the relevance matrix are updated at the same time during training using an update scheme similar to GLVQ and GRLVQ:

$$\begin{aligned} \mathbf{w}^{\mu+1}_{J} &= \mathbf{w}^{\mu}_{J} + 2\eta\frac{\Phi'(\tau_{\Lambda})d^{\mu}_{\Lambda,K}}{(d^{\mu}_{\Lambda,J} + d^{\mu}_{\Lambda,K})^2}\Lambda^{\mu}(\xi^{\mu} - \mathbf{w}^{\mu}_{J}), \\ \mathbf{w}^{\mu+1}_{K} &= \mathbf{w}^{\mu}_{K} - 2\eta\frac{\Phi'(\tau_{\Lambda})d^{\mu}_{\Lambda,J}}{(d^{\mu}_{\Lambda,J} + d^{\mu}_{\Lambda,K})^2}\Lambda^{\mu}(\xi^{\mu} - \mathbf{w}^{\mu}_{K}), \end{aligned} \tag{2.18}$$

where τ_{Λ}, is defined as in Equation (2.5), replacing d^{μ}_{T} by $d^{\mu}_{\Lambda,T}$. The update of the relevance matrix is given by Equation (2.19).

$$\begin{aligned} \Omega^{\mu+1} = \Omega^{\mu} - 2\epsilon\Phi'(\tau_{\Lambda}) \Bigg(&\frac{d^{\mu}_{\Lambda,K} \cdot \Omega^{\mu-1}(\xi^{\mu} - \mathbf{w}^{\mu}_{J})}{(d^{\mu}_{\Lambda,J} + d^{\mu}_{\Lambda,K})^2} \\ &- \frac{d^{\mu}_{\Lambda,J} \cdot \Omega^{\mu-1}(\xi^{\mu} - \mathbf{w}^{\mu}_{K})}{(d^{\mu}_{\Lambda,J} + d^{\mu}_{\Lambda,K})^2} \Bigg) \end{aligned} \tag{2.19}$$

The relevance matrix Ω was initialized using random samples drawn from on the interval $[-1, 1]$, i.e., $\Omega^0 \sim U_{M\times N}(-1, 1)$, and normalized such that the diagonal of $\Lambda^0 = \Omega^0(\Omega^0)^{\top}$ equals 1.

After training, the relevance matrix $\Lambda = \Omega\Omega^{\top}$ can be inspected to find the most influential features for the decisions taken by the classifier. In essence Ω forms a linear transformation $\Omega\Xi$ of the original input space to an $M \leq N$-dimensional space that allows for optimal separation of classes. Such linear transformations Ω are uniquely defined if and only if the matrix of training data $\mathbf{\Xi}_{\text{tr}}$ has rank N, i.e., it provides an N dimensional basis. Often $\mathbf{\Xi}_{\text{tr}}$ has a smaller rank due to correlated data or the availability of less data samples than there are data dimensions. As reasoned by Strickert et al. (2013), if $C = \Xi\Xi^{\top}$ has eigenvectors with eigenvalues zero, these correspond to directions in the feature space in which the data shows no variation. Adding arbitrary linear combinations of these vectors to the rows of Ω will not effect the projection $\Omega\Xi$. Hence, random effects can be observed in Ω, hence Λ obtained from GMLVQ. As outlined in Strickert et al. (2013), a transformation of Ω is required to eliminate the effect of randomness and find a unique solution for Ω and hence Λ. We apply the following post-processing:

$$\begin{aligned} \hat{\Omega} &= \Omega\Psi \\ \hat{\Lambda} &= \Psi^\top \Omega^\top \Omega \Psi \end{aligned} \tag{2.20}$$

where Ψ is constructed as:

$$\Psi = [\sum_{j=1}^{J} \mathbf{u}_j \mathbf{u}_j^\top] = [I - \sum_{j=J+1}^{N} \mathbf{u}_j \mathbf{u}_j^\top] \tag{2.21}$$

from the J non-vanishing eigenvectors $\mathbf{u}_1 \dots \mathbf{u}_J$ of $C = \Xi\Xi^\top$ given data Ξ.

2.3.2 Class-conditional means

Rather than training prototypes using an LVQ scheme, we also use the class conditional means as prototypes. This method can be considered a reference technique, and has e.g., been applied to Appraisal data by Scherer (1993), as is further outlined in Chapter 6. In a similar fashion to LVQ, unseen data is classified by returning the label of the prototype closest by. We chose to use squared Euclidean distance as distance measure, and implemented the method within our LVQ framework in Matlab (R2013a).

2.3.3 k-Nearest Neighbors

k-Nearest Neighbors (kNN) (see e.g., Duda et al. (2000)) is a technique in which all training data constitutes the knowledge of the classifier. Whenever a new data sample is provided, kNN searches for the k closest training samples and performs a majority vote. We used Euclidean distance as distance measure. We used the Matlab (R2013a) implementation of kNN from the Bioinformatics Toolbox (version 4.3).

2.3.4 Artificial Neural Network

Artificial Neural Networks (ANNs) (see e.g., Duda et al. (2000)) link input and output through a network consisting of layers of nodes. Each node is connected to all nodes of the previous layer and combines their input using a weighted sum and applies a transfer function. The input and output layers are directly observable, while the layers in between are not, therefore these are referred to as hidden layers. During training, errors on the output layer are propagated to previous layers by adjusting the weight factors. We used a three layer feed forward, back propagation network with the hyperbolic tangent transfer function on the hidden layer. In our

analyses we varied the number of nodes in the hidden layer (N_{hidden}). We used the implementation of ANN in the Neural Network Toolbox (version 8.0.1) of Matlab (R2013a).

2.3.5 Support Vector Machine

Support Vector Machine (SVM), introduced by Vapnik (1998), is a popular technique that optimizes the margin between data of two classes, i.e., it finds the optimal hyperplane separating data of two classes. It does so by first applying a transformation of data samples to a higher dimensional space using a kernel (we used linear and Radial Basis Function (RBF) kernels). Misclassification is penalized with a cost factor. Thereby, SVM minimizes:

$$\frac{1}{2}\|w\|^2 + C\sum_i \zeta_i \tag{2.22}$$

subject to:

$$y_i(w \cdot x_i + b) \geq 1 - \zeta_i, \text{ with } \zeta_i > 0, \tag{2.23}$$

where w is the hyperplane separating data samples $x_i = K(\xi_i)$ (already transformed using the kernel function K) with corresponding class membership indicators $y_i = \pm 1$, and cost (of misclassification) parameter C, which we varied in our analyses. In order to apply SVM to a multi-class problem, it requires an ensemble learning strategy. We have chosen to apply one-vs-rest in which for each class a classifier is created to distinguish it from all other classes, because it brings certain computational advantages while performing on par with other techniques (Rifkin and Klautau 2004). When classifying an unseen sample, all SVMs are applied and the class label corresponding to the SVM with largest margin (i.e., distance between the sample and the separating hyperplane) is returned. We used the LIBSVM (Chang and Lin 2011) implementation through an interface in Matlab (R2013a).

Material based on: Witoelar, A. W., Ghosh, A., de Vries, J. J. G., Hammer, B. and Biehl, M.: 2011, Window-based example selection in learning vector quantization, Neural Computation 22(11), 2924-2961.

Hoofdstuk 3

LEARNING DYNAMICS OF LEARNING VECTOR QUANTIZATION

Abstract

A variety of modifications has been employed to Learning Vector Quantization (LVQ) algorithms using either crisp *or* soft *windows for selection of data. Although these schemes have been shown in practice to improve performance, a theoretical study on the influence of windows has so far been limited. Here we rigorously analyse the influence of windows in a controlled environment of Gaussian mixtures in high dimensions. Concepts from statistical physics and the theory of on-line learning allow for an exact description of the training dynamics, yielding typical learning curves, convergence properties and achievable generalization abilities. We compare the performance and demonstrate the advantages of various algorithms, including LVQ 2.1, Generalized Learning Vector Quantization (GLVQ), Learning From Mistakes (LFM) and Robust Soft Learning Vector Quantization (RSLVQ). We find that the selection of the window parameter highly influences the learning curves, but surprisingly not the asymptotic performances of LVQ 2.1 and RSLVQ. Although the prototypes of LVQ 2.1 exhibit divergent behavior, the resulting decision boundary coincides with the optimal decision boundary thus yielding optimal generalization ability.*

3.1 Introduction

Out of many methods for classification, Learning Vector Quantization constitutes a family of learning algorithms for nearest prototype classification of potentially high dimensional data (Kohonen 2001). The intuitive approach and computational efficiency of LVQ classifiers have motivated its application in various disciplines, see e.g. Neural Networks Research Centre, Helsinki (2002). Prototypes in LVQ algorithms represent typical features within a data set using the same feature space instead of the black-box approach practiced in many other classifiers, e.g. feedforward neural networks or support vector machines. This makes them attractive to researchers outside the field of machine learning. Other advantages of LVQ algorithms are (1) they are easy to implement for multi-class classification tasks and (2) the algorithm complexity can be adjusted during training as required.

Numerous variants of the original LVQ prescriptions have been proposed towards achieving better performance, such as LVQ 2.1 (Kohonen 1990, Kohonen 2001), LVQ 3 (Kohonen 1990, Kohonen 2001), GLVQ (Hammer and Villmann 2002, Sato and Yamada 1995) and RSLVQ (Seo and Obermayer 2003). Common themes of these modifications include an additional parameter which controls the selection of data to which the system is adapted and variation of the magnitude of prototype updates. We refer to these in general as *window* schemes. In the limiting case of *hard* or *crisp* learning schemes, updates are restricted only to examples which fall into this window. For instance, LVQ 2.1 allows updates as long as the example is in the vicinity of the current decision boundary. Alternatively, learning schemes can implement a *soft* window, e.g. RSLVQ and GLVQ, which considers all examples but adapts the magnitude of the update according to their relative distances to the current decision boundary.

In general, the learning behavior of these strategies is not well understood. It is unclear how the convergence, stability and achievable generalization ability compare for the different strategies. Fortunately, methods from statistical physics and theory of on-line learning recently allowed a systematic investigation of very large systems in the so-called thermodynamic limit. This has been successfully applied in, among others, feedforward neural networks, perceptron training and principal component analysis (Biehl and Caticha 2003, Engel and van den Broeck 2001, Saad 1999). A similar approach to LVQ-type algorithms, e.g. LVQ 1, unsupervised Vector Quantization (VQ) and rank-based Neural Gas, was treated in Biehl et al. (2007) and Witoelar et al. (2008).

In this work, we closely examine the influence of window schemes for LVQ algorithms. Typical learning behavior is studied within a model situation of high dimensional Gaussian clusters and competing prototypes. From this analysis, we can observe typical learning curves and the convergence properties, i.e. the asymptotic behavior in the limit of an arbitrarily large number of examples.

Typically the window parameters are selected either heuristically or derived from prior knowledge of the data and kept fixed during training. The optimal parameter settings are chosen according to a computationally expensive validation procedure. It is also possible to treat the hyperparameters as dynamic properties during learning, e.g. by means of an annealing schedule (Seo and Obermayer 2006) or a gradient-based optimization method (Bengio 2000). Using the model described in this paper, one can investigate the optimality of the parameters for both fixed and dynamic settings in representative model situations.

3.2 Model

Throughout the paper, we study LVQ algorithms in a model situation: high dimensional data are generated from a mixture of M Gaussian clusters and presented to a system of two or three prototypes. We restrict ourselves to the analysis of isotropic and homogeneous clusters, i.e. each cluster σ generates only data with one of the class labels $y_\sigma \in \{1, 2, \ldots, N_c\}$ where N_c is the number of classes. Examples $\{\xi^\mu, y^\mu_\sigma\}$ with $\xi^\mu \in \mathbb{R}$ are drawn independently according to the probability density function

$$P(\xi) = \sum_{\sigma=1}^{M} p_\sigma P(\xi|\sigma) \;\text{ with }\; P(\xi|\sigma) = \frac{1}{(2\pi v_\sigma)^{N/2}} \exp\left[-\frac{1}{2v_\sigma}(\xi - \ell_\sigma \mathbf{B}_\sigma)^2\right] \qquad (3.1)$$

where p_σ are the cluster-wise prior probabilities and $\sum_\sigma p_\sigma = 1$. The components of vectors ξ^μ from cluster σ^μ are random numbers with mean vectors $\ell_\sigma \mathbf{B}_\sigma$ and variance v_σ. The unit vectors $\mathbf{B}_\sigma$ determine the orientation of cluster centers. Similar densities have been studied in Barkai et al. (1993), Biehl (1994), Biehl et al. (2007) and Meir (1995).

In this framework we formally exploit the thermodynamic limit $N \to \infty$ corresponding to very high dimensional data. This has simplifying consequences which will be present throughout the paper. Note that on random subspace projections, data from different clusters completely overlap and are not separable. The clusters become apparent only in the, at most, M-dimensional space spanned by vectors $\{\mathbf{B}_\sigma\}_{\sigma=1}^M$. The non-trivial goal is to identify this subspace from the N-dimensional data.

We bring attention to the readers on the scaling of the model. The anisotropy of this data distribution is very weak: while the mean of cluster σ, given by $\ell_\sigma \mathbf{B}_\sigma$, is a vector of length $\mathcal{O}(1)$, the average squared length of the data vectors $(\xi)^2$ is in the order $\mathcal{O}(N)$.

Obviously, this model is greatly simplified from practical situations. However it represents an ideal scenario to analyse the considered learning algorithms and Gaussian modeling of feature vectors which is a common technique in many practical scenarios. While more complex behaviors are expected in practical applications, the non-trivial effects already observed in this model will clearly influence the outcome under more general circumstances.

3.3 Algorithms

We shortly review LVQ algorithms and their corresponding window schemes. For the two-class model defined in Section 3.2, we define an LVQ system as a set of K

prototypes $W = \{\mathbf{w}_S, c_S\}_{S=1}^{K}$ with $\mathbf{w}_S \in \mathbb{R}$ and $c_S = \{1, 2, \ldots, N_c\}$. Classification is implemented through a nearest prototype scheme: novel examples will be assigned to the class of the closest prototype according to a dissimilarity measure. Here we restrict the measure to the squared Euclidean distance $d_S^\mu = (\xi^\mu - \mathbf{w}_S)^2$ for a given novel example ξ^μ. In this chapter, we investigate several LVQ prescriptions which include window schemes.

3.3.1 LVQ 2.1

The algorithm of LVQ 2.1 has been presented in Chapter 2 describing the update step and window used to restrain stability issues. However, this window is ineffective for very high dimensional data, as we obtain $\lim_{N\to\infty}(\xi^\mu - \mathbf{w}_S)^2 \approx (\xi^\mu)^2$ because $(\xi^\mu)^2 = \mathcal{O}(N)$ terms dominate the other $\mathcal{O}(1)$-terms, i.e. $(\mathbf{w}_S \cdot \xi^\mu)$ and $(\mathbf{w}_S^2)$. Consequently, this window definition does not work in very high dimensions, evidenced by

$$\lim_{N\to\infty} \min \left(\frac{\left(\xi^\mu - \mathbf{w}_T^{\mu-1}\right)^2}{\left(\xi^\mu - \mathbf{w}_S^{\mu-1}\right)^2}, \frac{\left(\xi^\mu - \mathbf{w}_S^{\mu-1}\right)^2}{\left(\xi^\mu - \mathbf{w}_T^{\mu-1}\right)^2} \right) = 1, \tag{3.2}$$

which implies that every example falls into the window. Therefore, in the following we implement the constraint

$$\left|(\xi^\mu - \mathbf{w}_T)^2 - (\xi^\mu - \mathbf{w}_S)^2\right| \leq k \ \min\left((\xi^\mu - \mathbf{w}_S)^2, (\xi^\mu - \mathbf{w}_T)^2\right) \tag{3.3}$$

where k is a small positive number. Note that the term $(\xi^\mu)^2 = \mathcal{O}(N)$ cancels out on the left hand side, while it dominates on the right hand side for $N \to \infty$. Thus, the right hand side becomes $k \cdot (\xi^\mu)^2$ and the condition is non-trivial only if $k = \mathcal{O}(1/N)$. We introduce the rescaled window parameter $\delta = k \cdot (\xi^\mu)^2 = \mathcal{O}(1)$ so that the window scheme is $-\delta \leq (d_T^\mu - d_S^\mu) \leq \delta$; δ is positive. We describe these rules as the following modulation function

$$f_S \quad = \quad \chi(c_S, y_\sigma^\mu) \sum_{T:c_T \neq c_S} \left(\Theta_{ST}^{-\delta} - \Theta_{ST}^{+\delta}\right) \prod_{U\neq S,T} \Theta_{SU}\Theta_{TU} \tag{3.4}$$

with $\chi(c_S, y_\sigma^\mu) = 1$ if $c_S = y_\sigma^\mu$ and $\chi(c_S, y_\sigma^\mu) = -1$ else. We use the shorthand notation $\Theta_{ji}^\delta \equiv \Theta(d_i^\mu - d_j^\mu - \delta)$, where $\Theta(x)$ is the Heaviside function $\Theta(x) = 1$ if $x > 0; 0$ else. We sum over prototypes $\{\mathbf{w}_T | c_T \neq c_S\}$ and terms $\left(\Theta_{ST}^{-\delta} - \Theta_{ST}^{+\delta}\right) = \Theta(d_T^\mu - d_S^\mu + \delta) - \Theta(d_T^\mu - d_S^\mu - \delta)$ enforce the window condition. The product term $\prod_{U\neq S,T} \Theta_{SU}\Theta_{TU}$ singles out instances where $\mathbf{w}_S$ and $\mathbf{w}_T$ are the two closest prototypes. This form of f_S allows for the analysis given in Section 3.4.

3.3.2 LFM-W

The performance of LFM, as described in Chapter 2 can be improved by including data selection of data using the window rule in Equation (3.3). We refer to this algorithm as LFM-W, represented by the modulation function

$$f_S = \begin{cases} \sum_{K:c_K \neq y_\sigma} \left(\Theta_{KS} - \Theta^{\delta}_{KS}\right) \psi(S,K) & \text{if } c_S = y^\mu_\sigma \\ \sum_{J:c_J = y_\sigma} \left(\Theta_{SJ} - \Theta^{\delta}_{SJ}\right) \psi(J,S) & \text{else.} \end{cases} \tag{3.5}$$

with $\Theta_{ji} \equiv \Theta(d^\mu_i - d^\mu_j)$ and $\psi(J,K) = \prod_{T:c_T=y_\sigma} \Theta_{JT} \prod_{U:c_U \neq y_\sigma} \Theta_{KU}$ which identifies cases with $\mathbf{w}_J$ being the correct winner and $\mathbf{w}_K$ being the incorrect winner: $\psi(J,K) = 1$ if this condition is fulfilled and $\psi(J,K) = 0$ else. Terms in parentheses single out misclassified examples which fall into the window.

3.3.3 GLVQ

GLVQ, as presented in Chapter 2, where $\Phi(\tau)$ can be used to define a window around the decision boundary. Here the usefulness of selecting a non-linear $\Phi(\tau)$ is shown. For instance, in Hammer and Villmann (2002) and Sato and Yamada (1995), the sigmoid function is chosen: $\Phi(\tau) = 1/(1+\exp(-\tau))$. The form of $\partial\Phi(\tau)/\partial\tau$, which has a single peak at $\tau = 0$, can be interpreted as a *soft* window around the decision boundary.

In the high dimensional limit, we notice that $(d^\mu_J + d^\mu_K)$ is dominated by $(\xi^\mu)^2$-terms and, effectively, becomes a constant $\mathcal{O}(N)$-term: $1/N(d^\mu_J + d^\mu_K) = 1 + \mathcal{O}(1/N)$. Therefore the denominator term in (2.5) becomes constant:

$$\lim_{N\to\infty} E = \lim_{N\to\infty} \sum_\mu \Phi\left(\frac{C}{d^\mu_J + d^\mu_K}(d^\mu_J - d^\mu_K)\right) = \sum_\mu \Phi\left(\frac{1}{v_G}\left(d^\mu_J - d^\mu_K\right)\right). \tag{3.6}$$

To obtain a non-zero argument, C must also be in the order $\mathcal{O}(N)$, and we rescale using $v_G = (d^\mu_J + d^\mu_K)/C = \mathcal{O}(1)$. The parameter v_G determines the softness of the window, provided that an appropriate non-linear $\Phi(\tau)$ is chosen. Note that GLVQ can be simplified to LVQ 2.1 without a window using the identity function $\Phi(\tau) = \tau$. The cost function in (2.5) becomes $E = \sum_\mu \left(d^\mu_J - d^\mu_K\right)/v_G$, where v_G could be set to 1 without changing its learning behavior. The modulation function is then reduced to $f_J = +1$, $f_K = -1$.

In this chapter, we choose the cumulative normal distribution

$$\Phi(\tau) = \int_{-\infty}^{\tau} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \tag{3.7}$$

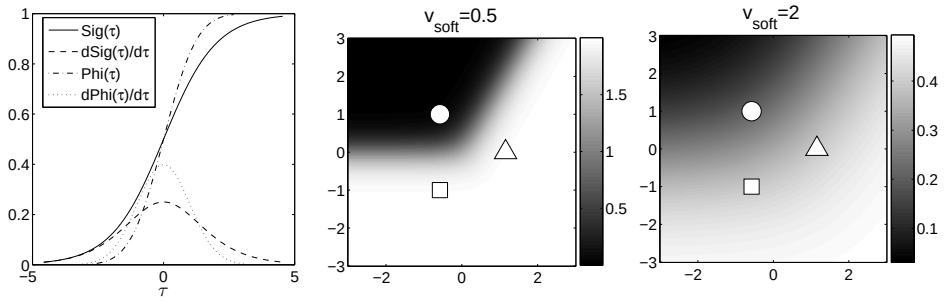

Figure 3.1: *Left panel: The form of the chosen $\Phi(\tau)$ in GLVQ, in comparison to the sigmoidal function Sig(τ). The derivatives produce a soft window. Middle and right panel: The RSLVQ modulation function f_S for class 1 ($\circ$) when presented with data from class 1. The figures display the difference between smaller v_{soft} (left) and larger v_{soft} (right).*

where $\partial\Phi(\tau)/\partial\tau = \phi(\tau) = \left(1/\sqrt{2\pi}\right)\exp\left(-\tau^2/2\right)$. Note that this form implements a Gaussian window similar to the sigmoidal cost described in Hammer and Villmann (2002) and Sato and Yamada (1995) and therefore produces a qualitatively similar behavior, see Figure 3.1 for the comparison.

Plugging in the form of (3.6), we obtain the learning rules

$$f_J = +\frac{2}{v_G}\,\phi\left(\frac{d_J - d_K}{v_G}\right), \qquad f_K = -\frac{2}{v_G}\,\phi\left(\frac{d_J - d_K}{v_G}\right) \tag{3.8}$$

We can write the modulation function as

$$f_S = \begin{cases} \displaystyle\sum_{K:c_K \neq y_\sigma} \left(\frac{2}{v_G}\phi\left(\frac{d_S - d_K}{v_G}\right)\right)\psi(S,K) & \text{if } c_S = y_\sigma^\mu \\ \displaystyle -\sum_{J:c_J = y_\sigma} \left(\frac{2}{v_G}\,\phi\left(\frac{d_J - d_S}{v_G}\right)\right)\psi(J,S) & \text{else.} \end{cases} \tag{3.9}$$

with $\psi(J,K) = \displaystyle\prod_{T:c_T = y_\sigma} \Theta_{JT} \prod_{U:c_U \neq y_\sigma} \Theta_{KU}$.

3.3.4 RSLVQ

In addition to the description of RSLVQ in Chapter 2, we can make the following observations: As v_{soft} becomes smaller, the updates become smaller for correctly classified examples and larger for incorrectly classified examples, see Figure 3.1.

Note than the limiting case of v_{soft} is particularly simple. The assignments of Equation (2.12) become hard assignments, i.e.

$$P_\sigma(S|\xi^\mu) = \begin{cases} 1, & \text{if } d_S^\mu = \min\limits_{\{j:c_j=\sigma^\mu\}} \left\{d_j^\mu\right\} \\ 0, & \text{else} \end{cases}, \; P(S|\xi^\mu) = \begin{cases} 1, & \text{if } d_S^\mu = \min\limits_{\{j\}} \left\{d_j^\mu\right\} \\ 0, & \text{else} \end{cases} \tag{3.10}$$

Plugging the above into (2.11), we obtain the learning rule for LFM, described in Section 3.3.2.

3.4 Analysis

In this section we describe the methods to analyse the learning dynamics in LVQ algorithms. Following the lines of the theory of on-line learning, see e.g. Biehl and Mietzner (1993), Biehl and Schwarze (1993), Engel and van den Broeck (2001) or Saad (1999), the system can be fully described in terms of a few characteristic quantities, so-called order parameters, in the thermodynamic limit $N \to \infty$. A suitable set of order parameters for the considered learning model is:

$$R_{S\sigma}^\mu = \mathbf{w}_S^\mu \cdot \mathbf{B}_\sigma \qquad Q_{ST}^\mu = \mathbf{w}_S^\mu \cdot \mathbf{w}_T^\mu. \tag{3.11}$$

Note that $R_{S\sigma}$ are the projections of prototype vectors $\mathbf{w}_S^\mu$ on the center vectors $\mathbf{B}_\sigma$ and Q_{ST}^μ correspond to the self- and cross- overlaps of the prototype vectors. From the generic update rule defined above, Equation (2.1), we can derive the following recursions in terms of the order parameters:

$$\begin{aligned} \frac{R_{S\sigma}^\mu - R_{S\sigma}^{\mu-1}}{1/N} &= \eta f_S\big(b_\sigma^\mu - R_{S\sigma}^{\mu-1}\big) \\ \frac{Q_{ST}^\mu - Q_{ST}^{\mu-1}}{1/N} &= \eta\big[f_T\big(h_S^\mu - Q_{ST}^{\mu-1}\big) + f_S\big(h_T^\mu - Q_{ST}^{\mu-1}\big)\big] + \eta^2 \frac{f_S f_T (\xi^\mu)^2}{N} + \mathcal{O}\left(\frac{1}{N}\right) \end{aligned} \tag{3.12}$$

where the input data vectors ξ^μ enter the system as their projections h_S^μ and b_σ^μ, defined as

$$h_S^\mu = \mathbf{w}_S^{\mu-1} \cdot \xi^\mu \qquad b_\sigma^\mu = \mathbf{B}_\sigma \cdot \xi^\mu. \tag{3.13}$$

In the limit $N \to \infty$, the $\mathcal{O}(1/N)$ term can be neglected and the order parameters *self average* (Reents and Urbanczik 1998) with respect to the random sequence of examples. This means that fluctuations of the order parameters vanish and the system dynamics can be described exactly in terms of their mean values. Also for $N \to \infty$, the rescaled quantity $\alpha \equiv \mu/N$ can be conceived as a continuous time variable. Accordingly, the dynamics can be described by a set of coupled Ordinary

Differential Equations (ODE) (Ghosh et al. 2006) after performing an average over the sequence of input data:

$$\begin{aligned}\frac{dR_{S\sigma}}{d\alpha} &= \eta\big(\langle b_\sigma f_S\rangle - \langle f_S\rangle R_{S\sigma}\big)\\ \frac{dQ_{ST}}{d\alpha} &= \eta\big(\langle h_S f_T\rangle - \langle f_T\rangle Q_{ST} + \langle h_T f_S\rangle - \langle f_S\rangle Q_{ST}\big) + \eta^2 \sum_\sigma p_\sigma \upsilon_\sigma \langle f_S f_T\rangle_\sigma\end{aligned} \tag{3.14}$$

where $\langle . \rangle$ and $\langle . \rangle_\sigma$ are the averages over the density $P(\xi)$ and $P(\xi|\sigma)$. To simplify the last term of Equation (3.14), we used

$$\lim_{N\to\infty} \langle f_S f_T \xi^2\rangle/N = \lim_{N\to\infty} \sum_\sigma p_\sigma(\upsilon_\sigma N + \ell^2)\langle f_S f_T\rangle_\sigma/N = \sum_\sigma p_\sigma \upsilon_\sigma \langle f_S f_T\rangle_\sigma.$$

In various sections in this paper, we investigate learning behaviors using small learning rates $\eta \to 0$ and neglect the η^2 terms in Equation (3.14). Non trivial behavior is only expected by taking the simultaneous limit $\eta \to 0, \alpha \to \infty$ and rescaling $\widetilde{\alpha} = \eta\alpha$ in Equation (3.14).

Exploiting the limit $N \to \infty$ once more, the quantities h_S^μ, b_σ^μ become correlated Gaussian quantities by means of the Central Limit Theorem. Therefore, they are fully specified by first and second moments, detailed in Appendix 3.A:

$$\begin{aligned}&\langle h_S^\mu\rangle_\sigma = \ell_\sigma R_{S\sigma}^{\mu-1}, \quad \langle b_\tau^\mu\rangle_\sigma = \ell_\sigma \delta_{\tau\sigma}, \quad \langle h_S^\mu h_T^\mu\rangle_\sigma - \langle h_S^\mu\rangle_\sigma\langle h_T^\mu\rangle_\sigma = \upsilon_\sigma Q_{ST}^{\mu-1}\\ &\langle b_\tau^\mu b_\rho^\mu\rangle_\sigma - \langle b_\tau^\mu\rangle_\sigma\langle b_\rho^\mu\rangle_\sigma = \upsilon_\sigma \mathbf{T}_{\tau\rho}, \quad \langle h_i^\mu b_\tau^\mu\rangle_\sigma - \langle h_i^\mu\rangle_\sigma\langle b_\tau^\mu\rangle_\sigma = \upsilon_\sigma R_{i\tau}^{\mu-1}.\end{aligned} \tag{3.15}$$

where S, T are prototype indices, τ, ρ, σ are cluster indices, δ is the Kronecker delta and $\mathbf{T}_{\tau\rho} \equiv \mathbf{B}_\tau \cdot \mathbf{B}_\rho$ is an overlap measure between clusters.

Thus, the above averages $\langle f_S\rangle$, $\langle h_T f_S\rangle$ and $\langle b_T f_S\rangle$ reduce to Gaussian integrations in $K + M$ dimensions and can be expressed in $\{R_{S\sigma}, Q_{ST}\}$, see Appendix 3.B. For various algorithms and a system with two competing prototypes, the averages can be calculated analytically. For three or more prototypes, the mathematical treatment becomes more involved and requires multiple numerical integrations.

Given the averages for a specific modulation function f_S, we obtain a closed set of ODE. Using initial conditions $\{R_{S\sigma}(0), Q_{ST}(0)\}$, we integrate this system for a given algorithm and obtain the evolution of order parameters in the course of training, $\{R_{S\sigma}(\alpha), Q_{ST}(\alpha)\}$. The generalization error ϵ_g, i.e. the probability of the closest prototype $\mathbf{w}_S$ carrying an incorrect label, is determined by considering the contribution from each cluster separately:

$$\epsilon_g = \sum_{\sigma=1}^{M} p_\sigma \epsilon_{g,\sigma} \;\text{ with }\; \epsilon_{g,\sigma} = \sum_{S:c_S\neq y_\sigma}^{K} \left\langle \prod_{T\neq S}^{K} \Theta_{ST} \right\rangle_\sigma, \tag{3.16}$$

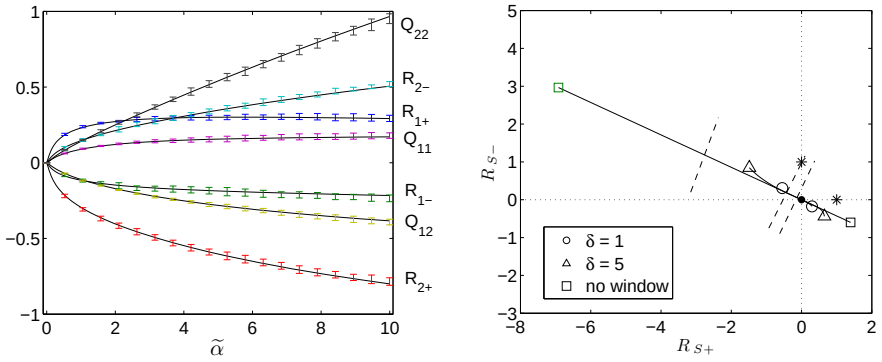

Figure 3.2: *Left panel: Evolution of the order parameters* $\{R_{S\sigma}, Q_{ST}\}$ *for LVQ 2.1 with* $K = 2$, $M = 2$, $\ell_1 = \ell_2 = 1, p_1 = 0.7, v_1 = v_2 = 1$ *and learning parameters* $\eta = 0.1$ *and* $\delta = 1$. *Solid lines represent* $\{R_{S\sigma}, Q_{ST}\}$ *obtained from the theoretical analysis, while bars represent the variance as produced by Monte Carlo simulations for* $N = 100$ *over 100 independent runs. Right panel: Influence of a window on LVQ 2.1 at learning time* $\alpha = 40$. *Prototypes are projected on the* $(\mathbf{B}_+, \mathbf{B}_-)$ *subspace for* $\delta = 1$ *(*$\circ$*),* $\delta = 5$ *(*$\triangle$*) and unrestricted LVQ 2.1 (*$\square$*). In the latter, one prototype strongly diverges. The resulting decision boundaries are indicated by chained lines. The origin is marked by (*$\cdot$*) and the cluster centers are marked by (*$*$*).*

which can be calculated from $\{R_{i\sigma}(\alpha), Q_{ij}(\alpha)\}$. For instance, for the simplest system with two clusters $\sigma = \{+, -\}$ and prototypes $\mathbf{w}_+$ and $\mathbf{w}_-$, the generalization error is written explicitly in terms of order parameters as

$$\epsilon_{g,\sigma} = \Phi\left(\frac{Q_{\sigma\sigma} - Q_{-\sigma,-\sigma} - 2\ell_\sigma(R_{\sigma,\sigma} - R_{-\sigma,\sigma})}{2\sqrt{v_\sigma}\sqrt{Q_{\sigma\sigma} - 2Q_{\sigma,-\sigma} + Q_{-\sigma,-\sigma}}}\right), \tag{3.17}$$

with $\Phi(x) = \int_{-\infty}^{x} dt \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)$, detailed in Appendix 3.D. The form of $\epsilon_{g,\sigma}$ for systems with more prototypes is more involved, and we refer the final result of the calculations to Appendix 3.D. We obtain the learning curve $\epsilon_g(\alpha)$ which quantifies the success of training. This method of analysis shows excellent agreement with Monte Carlo simulations of the learning system for dimensionality as low as $N = 100$, as demonstrated in Figure 3.2.

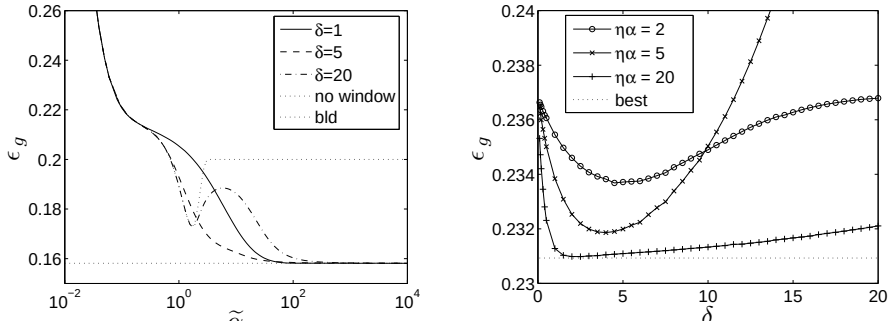

Figure 3.3: *Generalization error ϵ_g for LVQ 2.1 with $K = 2, M = 2, \ell = 1, p_+ = 0.8, p_- = 0.2$ and $\eta \to 0$. Left panel: ϵ_g vs $\widetilde{\alpha}$ using $\delta = 1, 5, 20$ and without a window. Note the logarithmic scaling on the horizontal axis. The asymptotic errors for all settings of δ converge at ${\epsilon_g}^{bld}$, indicated by the dotted line. Right panel: ϵ_g at fixed learning times $\widetilde{\alpha} = 2, 5$ and 20 as a function of δ.*

3.5 A simple case: two prototypes, two clusters

In this section we discuss in detail the results of the analysis for the simplest non-trivial problem: two-prototype LVQ 2.1, GLVQ, LFM-W and RSLVQ systems and $M = 2$ with one Gaussian cluster per class. The model data is given in Section 3.2. For simplicity, we denote the two clusters as $\sigma = \{+, -\}$ and without loss of generality can choose $\ell_+ = \ell_- = \ell$ and orthonormal $\mathbf{B}_\sigma$, i.e. $\mathbf{B}_i \cdot \mathbf{B}_j = 1$ if $i = j$; 0 else.

We place an emphasis on the asymptotic behavior in the limit $\alpha \to \infty$, i.e. the achieved performance for an arbitrarily large number of examples. The asymptotic generalization error $\epsilon_g(\infty)$ scales with the learning rate, analogous to minimizing a cost function in stochastic gradient descent procedures. For LVQ 2.1 and RSLVQ, the best achievable generalization error is obtained in the simultaneous limit of small learning rates $\eta \to 0$, $\alpha \to \infty$ and rescaling $\widetilde{\alpha} = \eta\alpha \to \infty$. However this limit is not meaningful for LFM, as will be explained later.

In this simple scenario, it is possible to exactly calculate the Best Linear Decision (BLD) boundaries by linear approximation of the Bayesian optimal decision boundary, see Biehl et al. (2004) for the calculations. We compare the results from each algorithm to the best linearly achievable error ${\epsilon_g}^{bld}$.

3.5.1 LVQ 2.1

We first examine two-prototype systems, i.e. $K = 2$. Figures 3.2 illustrate the evolution of order parameters under the influence of a window and the trajectories of the prototypes projected onto the $(\mathbf{B}_+, \mathbf{B}_-)$ subspace. Without additional constraints, LVQ 2.1 with two prototypes displays a strong divergent behavior in a system with unbalanced data, i.e. $p_+ \neq p_-$. The repulsion factor dominates for the prototype representing the weaker cluster, here $\mathbf{w}_2$. The order parameters associated with this prototype increase exponentially with $\widetilde{\alpha}$. As $\widetilde{\alpha} \to \infty$, $\mathbf{w}_2$ will be arbitrarily far away from the cluster centers and the asymptotic generalization error is trivial, $\epsilon_g(\infty) = \min(p_+, p_-)$.

Implementing the window scheme, $\mathbf{w}_2$ is repulsed until the data densities of both classes within the window become more balanced. Subsequently, the order parameters change with more balance between both prototypes. The repulsion factor still dominates its counterpart, therefore both prototypes still diverge, viz. $R_{S\sigma}$ for both prototypes display a linear change with $\widetilde{\alpha}$ at large $\widetilde{\alpha}$, but the decision boundary remains stable. Trivial classification is prevented, see the generalization error curves ϵ_g vs. $\widetilde{\alpha}$ in the left panel of Figure 3.3. Obviously, for smaller δ a considerable amount of data is filtered out and the initial learning stages slow down significantly. Meanwhile for large δ, ϵ_g becomes non-monotonic and converges more slowly.

Hence the performance at finite $\widetilde{\alpha}$ is dependent on δ, displayed in Figure 3.3, and parameter settings are highly critical in practical applications. Given learning time $\widetilde{\alpha}$, an optimal choice of fixed δ exists, which clearly depends on the properties of the data. With larger $\widetilde{\alpha}$, ϵ_g becomes less sensitive towards δ and the optimal setting of δ is smaller. Surprisingly, δ only influences the convergence speed while the non-trivial asymptotic generalization error $\epsilon_g(\infty)$ is insensitive to the choice of δ and equals the best achievable error ${\epsilon_g}^{bld}$ for each setting. This can be explained as follows. We can compare the asymptotic decision boundary to the BLD: the angle between them is equal to the angle between $(\mathbf{w}_1 - \mathbf{w}_2)$ and $(\mathbf{B}_+ - \mathbf{B}_-)$. This is calculated, using (3.11) and the orthonormality of $\mathbf{B}_+$ and $\mathbf{B}_-$, as

$$\varphi = \arccos\left(\frac{R_{1+} + R_{1-} - R_{2+} + R_{2-}}{\sqrt{2}\,(Q_{11} - 2Q_{12} + Q_{22})}\right), \tag{3.18}$$

which is found to be zero for large $\widetilde{\alpha}$. Hence, the decision boundary becomes parallel to the BLD and only its offset produces the difference between $\epsilon_g(\infty)$ and ${\epsilon_g}^{bld}$. In low dimensions, this offset oscillates around zero due to the window rule. In the thermodynamic limit, the fluctuations vanish and the LVQ 2.1 decision boundary coincides with the BLD.

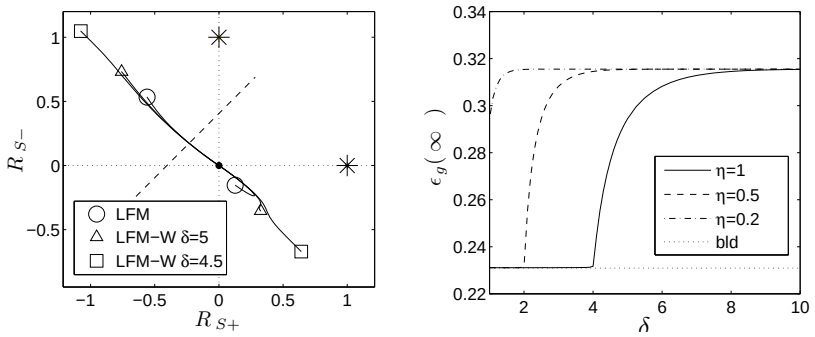

Figure 3.4: *LFM-W with* $p_+ = 0.6, \ell = 1, v_+ = v_- = 1$. *Left: Asymptotic prototype configuration for LFM and LFM-W* $\delta = 5$ *and* 4.5*, projected on* $\{\mathbf{B}_+, \mathbf{B}_-\}$. *Cluster centers* $\ell\mathbf{B}_+, \ell\mathbf{B}_-$ *are indicated by* $*$. *The projection of* $\mathbf{w}_1, \mathbf{w}_2$ *lie parallel to the symmetry axis* $\ell(\mathbf{B}_+ - \mathbf{B}-)$*, although they retain components orthogonal to the* $\{\mathbf{B}_+, \mathbf{B}_-\}$ *subspace. Right:* $\epsilon_g(\infty)$ *as a function of the window size* δ. *The lines correspond to learning rates* $\eta = 0.2, 0.5$ *and* 1.0.

3.5.2 LFM-W

The LFM scheme performs updates identical to LVQ 2.1 with the condition that the example is misclassified. A detailed investigation into the characteristics of $K = 2$ unrestricted LFM has been presented in Biehl et al. (2007). There, it was shown that LFM produces stable prototype configurations for finite learning rates η. The projection of the prototypes lies parallel to the symmetry axis $\ell(\mathbf{B}_+ - \mathbf{B}-)$, displayed in Figure 3.4. However the prototypes $\mathbf{w}_1$ and $\mathbf{w}_2$ retain components orthogonal to the two dimensional subspace spanned by the cluster centers, indicated by $Q_{ST} > R_{S+}R_{T+} + R_{S-}R_{T-}$ which implies

$$|\mathbf{w}_S|^2 > |R_{S+}\mathbf{B}_+ + R_{S-}\mathbf{B}_-|^2.$$

The asymptotic generalization error $\epsilon_g(\infty)$ is suboptimal and insensitive to η: the asymptotic decision boundary remains at an angle φ from the optimal hyperplane, c.f. Equation (3.18), independent of η. The Euclidean distance between prototypes is given by the quantity

$$\Delta_q = \sqrt{(\mathbf{w}_1 - \mathbf{w}_2)^2} = \sqrt{Q_{11} - 2Q_{12} + Q_{22}}\,, \tag{3.19}$$

which is found to be proportional to η for $\alpha \to \infty$. At $\eta \to 0$, $\Delta_q \to 0$ and the prototypes coincide, and this limit is not meaningful in LFM.

In this analysis, we observe that window schemes can dramatically improve performance of LFM. Using a window, the tilt of the decision boundary from the op-

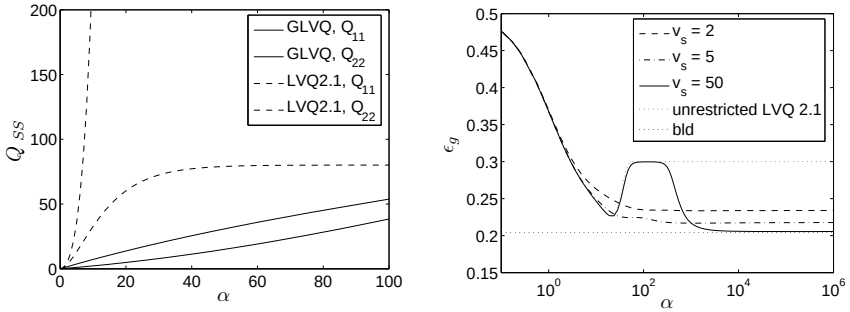

Figure 3.5: *Left panel: Q_{11} and Q_{22} for GLVQ (solid lines), compared to unrestricted LVQ 2.1 (dashed lines). The soft window of GLVQ slows down the repulsion of one prototype, but the prototypes remain divergent. Here $p_+ = 0.7, \ell = 1, v_+ = 2, v_- = 5$, $\eta = 0.25$. Right panel: Learning curves ϵ_g vs. α for softness $v_G = 2, 5$ and 50, note the logarithmic horizontal axis. The learning rates are maintained at $\eta/v_G = 0.1$. Large v_G produces better asymptotic generalization error, but may exhibit non-monotonic behavior and require very long learning times.*

timal hyperplane, i.e φ in (3.18), is reduced, resulting in lower $\epsilon_g(\infty)$. We observe that $\epsilon_g(\infty)$ decreases along with reducing δ, displayed in the right panel of Figure 3.4. However, a critical window size δ_c exists where the LFM unexpectedly becomes divergent and no stationary state exists. Smaller windows filter examples which produce more repulsion in the orientation of the cluster centers, and we observe asymptotically larger Δ_q as δ decreases. This is clearly observed in Figure 3.4. Given a sufficiently small δ, it is possible that the repulsion factor entirely outweighs the attractive factor. At $\delta < \delta_c$, it performs similar to LVQ 2.1: the angle φ becomes zero and $\epsilon_g(\infty)$ is close to the best achievable error.

Unlike the unrestricted case, the learning rate η can influence the asymptotic performance. The learning rate and window size are indirectly related, as shown in the right panel of Figure 3.4. For example, learning with small learning rates requires smaller windows to achieve optimal asymptotic error. Note that the influence of the window size depends heavily on the structure of the data. For various data models, efficient window settings may only exist on a very limited range and window schemes may be ineffective to improve generalization performance while still maintaining stability.

3.5.3 GLVQ

Apart from the influence of v_G to the overall learning rate, small v_G corresponds to a sharp peak around the decision boundary while large v_G corresponds to a very

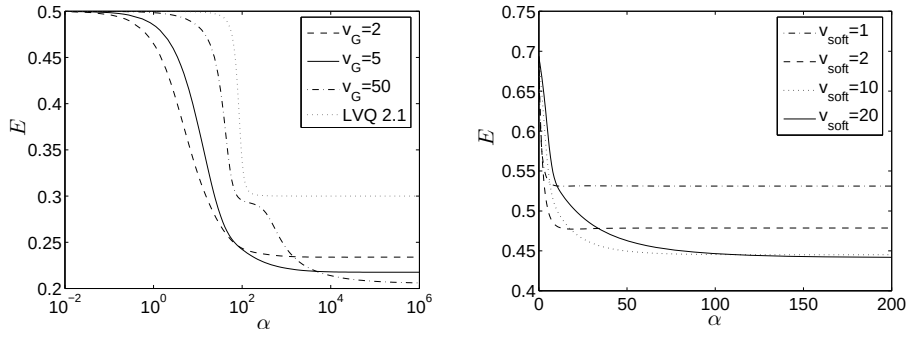

Figure 3.6: $p_+ = 0.7, \ell = 1, v_+ = 1, v_- = 1$. *The cost functions for GLVQ with* $\eta \to 0$ *(left panel) and RSLVQ with* $\eta/v_{soft} = 1$ *(right panel) decrease monotonically, corresponding to a stochastic gradient descent.*

large window. Figure 3.5 displays the prototype lengths while using GLVQ: the soft window slows down the strong repulsion of the prototype of the weaker cluster, as opposed to unrestricted LVQ 2.1. While both prototypes still diverge because the cost function at $N \to \infty$ is not bounded, c.f. (3.6), the asymptotic ϵ_g remains non-trivial, see Figure 3.5.

Note that v_G directly relates to the overall learning rate η/v_G, refer to Equation (3.9), which influences the level of noise in stochastic gradient procedures. We compare results with respect to v_G, while maintaining at equal overall learning rate by keeping η/v_G constant, in Figure 3.5. Performance deteriorates at smaller v_G, where training slows down at intermediate stages and converges at a higher error. However, very large v_G allows strong repulsion of the weaker prototype which results in non-monotonic ϵ_g and long learning convergence times. Surprisingly, the soft GLVQ window is outperformed by the simple hard or crisp window of LVQ 2.1. This is caused by the long tail of the modulation function which sums up into a large repulsion, whereas in the crisp window, only data near the decision boundary are considered.

Figure 3.6 displays the cost function during learning. In the initial learning stages, the minimization of the cost function E leads to fast decrease of ϵ_g. However, while the cost function continues to decrease monotonically, ϵ_g behaves non-monotonically. While many techniques are developed to improve minimization procedures of E, it is important to evaluate the choice of E and its correlation to the desired generalization performance.

3.5.4 RSLVQ

Finally in this section, we study the influence of the softness parameter v_{soft} in the RSLVQ algorithm. Note that in Seo and Obermayer (2003), the learning rate η and softness parameter v_{soft} are treated independently using separate annealing schedules. In this section, we assume η decreases proportionally with v_{soft}, i.e. a fixed overall learning rate η/v_{soft} is maintained.

We first investigate model scenarios with equal variance clusters $v_+ = v_-$ and unbalanced data $p_+ \neq p_-$. We observe the influence of v_{soft} on the learning curves, displayed on the left panel of Figure 3.7. The generalization error curve depends on v_{soft}: at large v_{soft}, ϵ_g may exhibit non-monotonic behavior, reminiscent of LVQ 2.1. Because of this behavior, the learning process may require long learning times before reaching the asymptotic configuration. This is an important consideration for practical applications which often uses early stopping strategies to avoid overtraining. Meanwhile, the algorithm minimizes the cost function E in (2.10) monotonically, see Figure 3.6. Thus, the decrease in E does not always result in a decrease of ϵ_g.

A major advantage of the RSLVQ algorithm is the convergence of prototypes, i.e. a stationary configuration of order parameters exists for finite v_{soft}. The asymptotic configuration of prototypes are displayed in Figure 3.8. At $\widetilde{\alpha} \to \infty$, the softness parameter controls only the distance between the two prototypes: Δ_q as defined in Equation (3.19), decreases linearly with v_{soft}. Note that under the conditions $p_+ = 0.5, v_{\text{soft}} = v_+ = v_-$ and initialization of prototypes on the symmetry axis, each prototype is located at its corresponding cluster center, i.e. the RSLVQ mixture model matches exactly to the actual input density.

Figures 3.7 compare the asymptotic errors in the case of $\eta/v_{\text{soft}} = 1$ (left panel) and small learning rates $\eta/v_{\text{soft}} \to 0$ (right panel). In the former case, performance improves with large v_{soft}: at small v_{soft}, the system converges at high ϵ_g similar to LFM, while at larger v_{soft}, it approaches the best linear decision. Meanwhile, at small learning rates, the asymptotic error becomes independent to v_{soft}. Therefore, given sufficiently small learning rates, RSLVQ becomes robust wrt. its softness parameter.

In the equal variance scenario, the asymptotic decision boundary always converges to the best linear decision boundary for all settings of $\{p_+, p_-\}$ and RSLVQ outperforms both LFM and LVQ 2.1, as it provides robustness, stability and low generalization error.

On the other hand, a scenario with unequal class variances presents an interesting case where RSLVQ with global v_{soft} fails to match the model. RSLVQ remains robust, i.e. the decision boundary converges to identical configurations for all settings of v_{soft}, see Figure 3.8. However, the asymptotic results are suboptimal. While

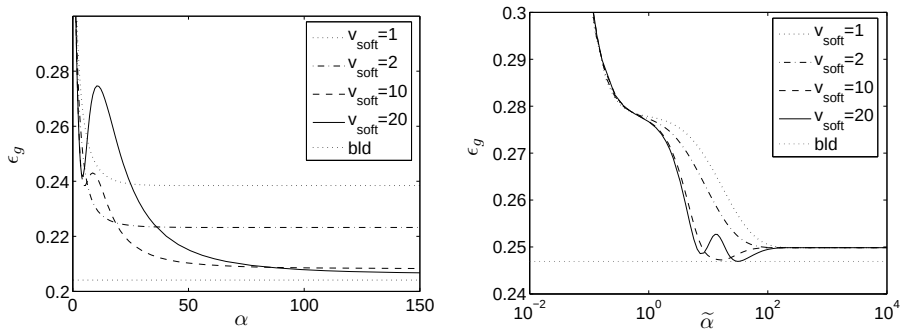

Figure 3.7: *Learning curves ϵ_g for RSLVQ using softness parameter $v_{soft} = 1, 2, 10$ and 20. Left: $p_+ = 0.7$ and equal variance $v_+ = v_- = 1$ with fixed overall learning rate $\eta/v_{soft} = 1$. Right: $p_+ = 0.6$ and unequal variance $v_+ = 1, v_- = 4$ with $\eta/v_{soft} \to 0$. The asymptotic errors is independent of v_{soft} at small learning rates, but at a suboptimal value. Note the the logarithmic scale of $\widetilde{\alpha}$.*

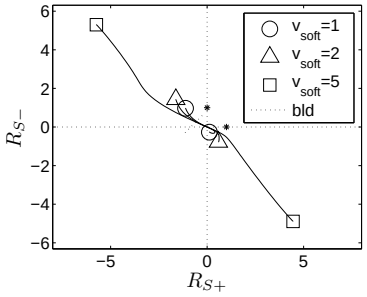

Figure 3.8: *Trajectories of prototypes of the system in the left panel of Figure 3.7. Prototypes are projected on the space Span$(\mathbf{B}_+, \mathbf{B}_-)$ for $v_{soft} = 1$ (circle), 2 (triangle) and 5 (square).*

RSLVQ is insensitive to the priors of the clusters, its performance wrt. the best achievable error is sensitive to the cluster variances, e.g. at highly unbalanced σ_+/σ_-, RSLVQ generalizes poorly and is outperformed by the simpler LVQ 2.1. In practical applications, v_{soft} may be set locally for each prototype to accomodate such scenarios, but this case cannot be treated along the lines of the present analysis in a straightforward way.

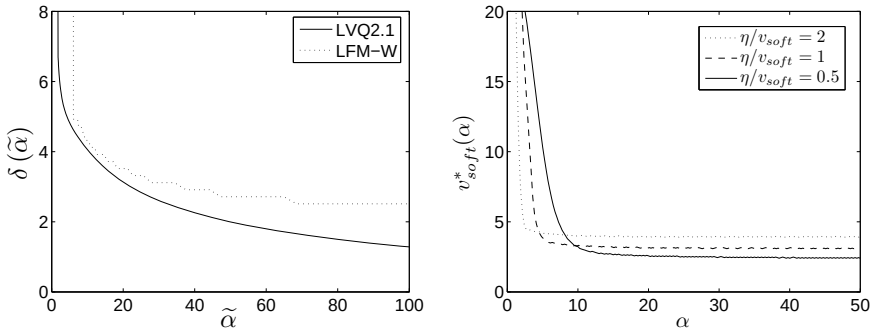

Figure 3.9: *Optimal window schedule $\delta(\alpha)$ for LVQ 2.1 and LFM, obtained by formally minimizing $d\epsilon_g/d\alpha$ with respect to $\delta(\alpha)$. Right panel: optimal softness parameter for RSLVQ with fixed $\eta/v_{soft} = 2, 1$ and 0.5.*

3.6 Optimal window schedules

We have observed in Sections 3.5.1 and 3.5.2 the learning curves and asymptotics of LVQ 2.1 and LFM-W wrt. fixed window parameters. In this section we treat the window parameter as dynamic properties during learning, viz. $\delta(\alpha)$. Although small windows allow optimal $\epsilon_g(\alpha \to \infty)$, their obvious disadvantage is their slower initial learning and convergence speed. This suggests that dynamic performance can be improved by adjusting the window along with the number of examples presented.

We calculate the locally optimal $\delta^*(\alpha)$-schedule by formally minimizing $d\epsilon_g(\alpha)/d\alpha$ with respect to δ using the knowledge of the input density and finding the condition

$$\delta^*(\alpha) \arg\min_{\delta} \left(\mathbf{u}(\alpha) \cdot \frac{d\mathbf{O}(\alpha)}{d\alpha} \right) = 0 \ \text{ with } \ \mathbf{u}(\alpha) = \sum_{\sigma=1}^{M} p_\sigma \frac{d\epsilon_{g,\sigma}(\alpha)}{d\mathbf{O}} \tag{3.20}$$

where we use the shorthand $\mathbf{O}$ for the set of order parameters. For a system with two prototypes $\{\mathbf{w}_+.\mathbf{w}_-\}$ and two clusters $\sigma = \{+,-\}$, $\mathbf{O} = \{R_{++}, R_{+-}, R_{-+}, R_{--}, Q_{++},$

$Q_{+-}, Q_{--}\}^T$ and derivating from (3.17), we obtain

$$\frac{d\epsilon_{g_\sigma}(\alpha)}{d\mathbf{O}} = \frac{1}{2\sqrt{v_\sigma}\Delta_q}\,\phi\left(\frac{Z_\sigma}{2\sqrt{v_\sigma}\Delta_q}\right).\mathbf{A}_\sigma \text{ with}$$

$$\mathbf{A}_+ = \begin{bmatrix} -2\ell \\ 0 \\ +2\ell \\ 0 \\ 1 - Z_+/(2\Delta_q^2) \\ Z_+/\Delta_q^2 \\ -1 - Z_+/(2\Delta_q^2) \end{bmatrix}, \quad \mathbf{A}_- = \begin{bmatrix} 0 \\ +2\ell \\ 0 \\ -2\ell \\ -1 - Z_-/(2\Delta_q^2) \\ Z_-/\Delta_q^2 \\ 1 - Z_-/(2\Delta_q^2) \end{bmatrix}$$

with $Z_\sigma = Q_{\sigma\sigma} - Q_{-\sigma,-\sigma} - 2\ell(R_{\sigma\sigma} - R_{-\sigma\sigma})$ and Δ_q defined in Equation (3.19), see Appendix 3.D for the calculations.

We plug in $d\mathbf{O}/d\alpha$ for the corresponding algorithm and numerically calculate $\delta^*(\alpha)$ from Equation (3.20) at each learning step. We find that the learning curve is improved with initially large δ which is decreased during training, following the curve in Figure 3.9. This suggests that practical schedules with gradual reduction of window sizes are indeed suitable for this particular learning problem.

While this approach locally minimizes generalization error, this strategy does not always lead to minimization of ϵ_g over a time span, i.e. a globally optimal schedule, which requires calculations along the lines of variational optimization, see e.g. Biehl (1994) or Saad and Rattray (1997), for its application of optimal learning rates in multilayered neural networks. Obviously, a priori knowledge of the input density is not available in practical situations. Nevertheless, this minimization technique provides an upper bound of the achievable performance of the learning scheme for a given model.

Figure 3.7 displays that although large v_{soft} for RSLVQ allows for a faster initial learning, it also can yield non-monotonic learning curves. We can avoid the non-monotonic behavior and maximize the decrease of ϵ_g by applying a variational approach analogous to (3.20) in order to calculate the locally optimal softness parameter schedule $v^*_{\text{soft}}(\alpha)$. While fixing the value of η/v_{soft}, we produce the locally optimal softness schedule $v^*_{\text{soft}}(\alpha)$ in Figure 3.9, where $v^*_{\text{soft}}(\alpha)$ is initially large and decreases to saturate at a constant value. Note that this value depends on the learning rate, e.g. it decreases with η/v_{soft}. In calculations with $\eta \to 0$, we obtain the limit $v^*_{\text{soft}}(\infty) \to 0$, which is the clearly suboptimal LFM. Therefore an analysis of optimal RSLVQ schedule requires $\eta > 0$.

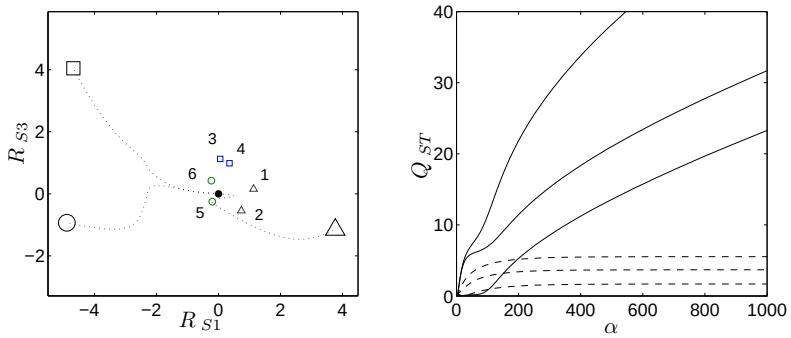

Figure 3.10: *Left panel: Snapshot at $\alpha = 50$ of an LVQ 2.1 system, $\delta = 1$ with $K = 3$ and $M = 6$ randomly generated isotropic clusters projected on the $(\mathbf{B}_1, \mathbf{B}_3)$ subspace. The solid dot marks the initial position of all prototypes and solid lines mark the trajectories of the prototypes. Right panel: $p_\triangle = 0.5, p_\square = 0.3, p_\circ = 0.2$. Solid lines represent, from bottom to top, prototype vector lengths Q_{11}, Q_{22}, Q_{33} for LVQ 2.1 $\delta = 10$. Dashed lines represent the result for RSLVQ $v_{soft} = 2$.*

3.7 Three-prototype systems

In this section we look at more generic analyses of LVQ algorithms by extending the previous systems to $K = 3$ prototypes and M clusters, requiring a much larger set of order parameters. This allows an initial study on two important issues concerning practical applications of LVQ: multi-class problems and the use of multiple prototypes within a class.

We first look at multi-class problems with $N_c = 3$ classes, an example is shown in Figure 3.10 for LVQ 2.1 with $M = 6$ clusters selected with random variances and random deviation from the original class centers. The clusters are separable only in M out of N dimensions. In all our observations, we find that the behaviors of $K = 3$ systems are qualitatively similar to $K = 2$ systems. For LVQ 2.1, the learning curves vary according to the window sizes, but its asymptotic generalization error is independent of δ. Due to the presence of other prototypes, the repulsion on a weaker class prototype are reduced. However, the prototypes remain divergent, e.g. Figure 3.10. Meanwhile for LFM-W, the asymptotic performance is sensitive to δ whose range of effective window sizes depend strongly on the learning parameters. For GLVQ, the prototypes are divergent with a higher asymptotic error than LVQ 2.1, and thus it performs poorly. Finally, for RSLVQ, the prototypes remain stable and the asymptotic generalization performance is robust wrt. settings of v_{soft}, but it is

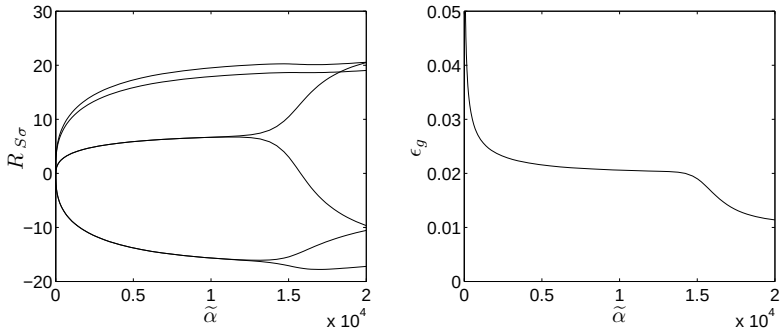

Figure 3.11: *Unspecialized phase induces long learning plateaus, shown with LFM-W* $K = 3, c_K = \{\pm 1\}$ *and input density* $M = 6$ *and* $N_c = 2, c_\sigma = \{\pm 1\}$. *Left panel: Several order parameters display a specialization phase between prototypes of same class. Right panel: Generalization error.*

outperformed by LVQ 2.1. Hence, the results are consistent with the $K = 2$ system and the preceding analysis is valid qualitatively to, at least, systems of M clusters and one prototype per class within the model restrictions.

To allow more complex decision boundaries, practical LVQ applications frequently employ several prototypes within a class. We investigate a two-class system $N_c = 2, y_\sigma = \{+, -\}$ using $K = 3$ prototypes with labels $c_S = \{+, +, -\}$ and observe the non-trivial interaction between similarly labeled prototypes, here $\mathbf{w}_1$ and $\mathbf{w}_2$. While prototypes of different classes immediately separate in the initial training phase, prototypes of the same class remain identical in the M dimensional space, see Figure 3.11. The latter prototypes differ only in dimensions which are not related for classification and produce a suboptimal decision boundary. This may proceed for a long learning period before these prototypes begin to *specialize*, i.e. each prototype produces a bigger overlap $R_{S\sigma}$ with a distinct group of clusters. The specialization phase produces a sudden decrease of ϵ_g, displayed in the right panel of Figure 3.11. This phenomenon is highly reminiscent of symmetry breaking effects observed in unsupervised learning, such as Winner-Takes-All VQ (Biehl 1994, Witoelar et al. 2008) or multilayer neural networks (Saad and Solla 1995).

Learning parameters highly influence the nature of the transition, e.g. large learning rates and smaller windows prolong the unspecialized phase, and therefore they are critical to the success of learning. Symmetry breaking may require exceedingly long learning times, resulting in learning plateaus which dominate the training process

and present a challenge in practical situations with very high dimensional data. In more extreme circumstances, the system may not escape the unspecialized state at all and the optimal classification cannot be obtained. Details of the symmetry breaking properties wrt. parameters will be investigated in following publications.

3.8 Conclusion

We have investigated the learning behavior of LVQ 2.1, GLVQ, LFM-W and RSLVQ using window schemes which work in high dimensions. The analysis is based on the theory of on-line learning on a model of high dimensional isotropic clusters. Our findings demonstrate that the selection of proper window sizes is critical to efficient learning for all algorithms. Given more available data and allowance for costly learning times, parameter selection becomes much less important.

Our analysis demonstrates the influence of windows on the learning curves and the advantages and drawbacks of each algorithm within the model scenarios. A summary is described in Table 3.1. Asymptotically, LVQ 2.1 achieves optimal performance in all scenarios, but stability remains an issue in terms of diverging prototypes. LFM-W shows a remarkable improvement in performance over LFM. Unfortunately, the introduction of a window may also influence its stability, and therefore it is highly parameter sensitive, i.e., only a narrow range of window size can improve the overall performance. GLVQ behaves similarly to LVQ 2.1. While GLVQ reduces the initial strong overshooting of LVQ 2.1, the prototypes remain divergent and GLVQ produces higher generalization errors or long convergence times. RSLVQ attempts to combine the advantages of both LFM and LVQ 2.1 by providing both stability and optimal performance. However, an important issue of RSLVQ lies on its approximation of the data structure, e.g., it performs well when the actual input density are isotropic Gaussian clusters with equal variance. If the assumptions depart from the input density, the results become suboptimal and RSLVQ can even be outperformed by the simpler LVQ 2.1 and LFM-W. In all scenarios, RSLVQ displays robustness of its classification behavior with respect to the softness parameter, given sufficiently low learning rates.

This analysis also allows a formal optimization of the window size during learning to ensure fast convergence. While in general, various window sizes for LVQ 2.1 produce equal asymptotic errors, initial window sizes should be chosen large for faster convergence speed and decreased in the course of learning. Similarly, an optimal schedule for RSLVQ points to a gradual decrease of the softness parameter to a particular saturation value, which agrees well with many practical scheduling schemes. However, locally optimal schedules do not always lead to the globally optimal sche-

Table 3.1: *Asymptotic properties of LVQ algorithms.*

	LVQ 2.1	LFM-W	GLVQ	RSLVQ
Stability	divergent	convergent*	divergent	convergent
Sensitivity wrt. parameters	robust	dependent	dependent	robust
Gen. ability	optimal	suboptimal	suboptimal	suboptimal

* under the condition that δ is larger than critical window size δ_c.

dules see, for instance, Saad and Rattray (1997). In further work, we will develop efficient dynamic parameter adaptations, i.e., optimal window schedules during on-line training along the lines of variational optimization.

We show that the analysis remains valid for multi-class systems and arbitrary number of isotropic clusters. Additionally, using multiple prototype assignments within a class, we already observe the presence of learning plateaus in this highly simplified scenario. These phenomena carry on and could dominate the training process in any practical situations with high degrees of freedom. Further investigations of more complex network architectures and non-trivial input distributions may also yield additional phenomena, e.g., competing stationary states of the system, and provide further insights to general LVQ behaviors.

3.A Statistics of the projections

For convenience, we combine the projections $h_S = \mathbf{w}_S \cdot \xi$ and $b_\sigma = \mathbf{B}_\sigma \cdot \xi$ defined in (3.13) into a D-dimensional vector, where $D = K + M$, as

$$\mathbf{x} = \left(\begin{array}{cccccccc} h_1^\mu & h_2^\mu & \dots & h_K^\mu & b_1^\mu & b_2^\mu & \dots & b_M^\mu \end{array} \right)^T \tag{3.21}$$

In our analysis of on-line learning, we assume that ξ is statistically independent from $\mathbf{w}_S$, because ξ^μ is uncorrelated to all previous data and $\mathbf{w}_S^{\mu-1}$. Therefore we observe that h_S and b_σ become correlated Gaussian random quantities following the Central Limit Theorem and can be fully described by their first and second moments, i.e. its conditional averages $\mu_\sigma = \langle \mathbf{x} \rangle_\sigma$ and conditional covariance matrix $C_\sigma = \langle \mathbf{x} \cdot \mathbf{x}^T \rangle_\sigma$. We compute these averages in the following.

3.A.1 First order statistics

We compute the averages of the components of $\mathbf{x}$ as follows:

$$\langle h_i \rangle_\sigma = \int_{\mathbb{R}^N} \xi \cdot \mathbf{w}_i p(\xi|\sigma) d\xi = \mathbf{w}_i \cdot \int_{\mathbb{R}^N} \xi p(\xi|\sigma) d\xi = \mathbf{w}_i \cdot \ell_\sigma \mathbf{B}_\sigma = \ell_\sigma R_{i\sigma} \quad (3.22)$$

$$\langle b_\tau \rangle_\sigma = \int_{\mathbb{R}^N} \xi \cdot \mathbf{B}_\tau p(\xi|\sigma) d\xi = \mathbf{B}_\tau \cdot \int_{\mathbb{R}^N} \xi p(\xi|\sigma) d\xi = \mathbf{B}_\tau \cdot \ell_\sigma \mathbf{B}_\sigma = \ell_\sigma T_{\tau\sigma} \quad (3.23)$$

with $T_{\tau\sigma} = \mathbf{B}_\tau \cdot \mathbf{B}_\sigma$. To a large extent, we utilize orthonormal cluster center vectors, i.e. $\mathbf{B}_\tau \cdot \mathbf{B}_\sigma = \delta_{\tau\sigma}$ where δ is the Kronecker delta. The conditional first order moments $\mu_\sigma = \langle \mathbf{x} \rangle_\sigma$ can be expressed in terms of order parameters as

$$\mu = \ell_\sigma \begin{pmatrix} R_{1\sigma} & R_{2\sigma} & \dots & R_{K\sigma} & T_{1\sigma} & T_{2\sigma} & \dots & T_{M\sigma} \end{pmatrix}^T \quad (3.24)$$

3.A.2 Second order statistics

To compute the conditional variance $\langle \mathbf{x}_n \mathbf{x}_m \rangle_\sigma - \langle \mathbf{x}_n \rangle_\sigma \langle \mathbf{x}_m \rangle_\sigma$ we first look at the average

$$\begin{aligned}
\langle h_i h_j \rangle_\sigma &= \left\langle \left(\sum_{k=1}^N (\mathbf{w}_i)_k (\xi)_k \right) \left(\sum_{l=1}^N (\mathbf{w}_j)_l (\xi)_l \right) \right\rangle_\sigma \\
&= \left\langle \sum_{k=1}^N (\mathbf{w}_i)_k (\mathbf{w}_j)_k (\xi)_k (\xi)_k + \sum_{k=1}^N \sum_{l=1, l\neq k}^N (\mathbf{w}_i)_k (\mathbf{w}_j)_l (\xi)_k (\xi)_l \right\rangle_\sigma \\
&= \sum_{k=1}^N (\mathbf{w}_i)_k (\mathbf{w}_j)_k \langle (\xi)_k (\xi)_k \rangle_\sigma + \sum_{k=1}^N \sum_{l=1, l\neq k}^N (\mathbf{w}_i)_k (\mathbf{w}_j)_l \langle (\xi)_k (\xi)_l \rangle_\sigma \\
&= \sum_{k=1}^N (\mathbf{w}_i)_k (\mathbf{w}_j)_k \left(\upsilon_\sigma + \ell_\sigma^2 (\mathbf{B}_\sigma)_k (\mathbf{B}_\sigma)_k \right) + \sum_{k=1}^N \sum_{l=1, l\neq k}^N (\mathbf{w}_i)_k (\mathbf{w}_j)_l \ell_\sigma^2 (\mathbf{B}_\sigma)_k (\mathbf{B}_\sigma)_l \\
&= \upsilon_\sigma \sum_{k=1}^N (\mathbf{w}_i)_k (\mathbf{w}_j)_k + \ell_\sigma^2 \sum_{k=1}^N \sum_{l=1}^N (\mathbf{w}_i)_k (\mathbf{w}_j)_l (\mathbf{B}_\sigma)_k (\mathbf{B}_\sigma)_l \\
&= \upsilon_\sigma \mathbf{w}_i \cdot \mathbf{w}_j + \ell_\sigma^2 (\mathbf{w}_i \cdot \mathbf{B}_\sigma)(\mathbf{w}_j \cdot \mathbf{B}_\sigma) = \upsilon_\sigma Q_{ij} + \ell_\sigma^2 R_{i\sigma} R_{j\sigma} \quad (3.25)
\end{aligned}$$

Here we exploit the following

$$\langle (\xi)_k (\xi)_k \rangle_\sigma = \upsilon_\sigma + \langle (\xi)_k \rangle_\sigma \langle (\xi)_k \rangle_\sigma = \upsilon_\sigma + \ell_\sigma^2 (\mathbf{B}_\sigma)_k (\mathbf{B}_\sigma)_k$$
$$\text{and} \quad \langle (\xi)_k (\xi)_l \rangle_\sigma = \langle (\xi)_k \rangle_\sigma \langle (\xi)_l \rangle_\sigma = \ell_\sigma^2 (\mathbf{B}_\sigma)_k (\mathbf{B}_\sigma)_l$$

Hence we obtain the conditional second order moment, from Eqs. (3.25) and (3.22),

$$\langle h_i h_j \rangle_\sigma - \langle h_i \rangle_\sigma \langle h_j \rangle_\sigma = \upsilon_\sigma Q_{ij} + \ell_\sigma^2 R_{i\sigma} R_{j\sigma} - \ell_\sigma R_{i\sigma} \ell_\sigma R_{j\sigma} = \upsilon_\sigma Q_{ij} \quad (3.26)$$

Analogously, we get the second order statistics of b and the covariance as follows:

$$\langle b_\tau b_\rho \rangle_\sigma - \langle b_\tau \rangle_\sigma \langle b_\rho \rangle_\sigma = v_\sigma T_{\tau\rho} + \ell_\sigma^2 T_{\tau\sigma} T_{\rho\sigma} - \ell_\sigma T_{\tau\sigma} \ell_\sigma T_{\rho\sigma} = v_\sigma T_{\tau\rho} \quad (3.27)$$

$$\langle h_i b_\tau \rangle_\sigma - \langle h_i \rangle_\sigma \langle b_\tau \rangle_\sigma = v_\sigma R_{i\tau} + \ell_\sigma^2 R_{i\sigma} T_{\tau\sigma} - \ell_\sigma R_{i\sigma} \ell_\sigma T_{\tau\sigma} = v_\sigma R_{i\tau} \quad (3.28)$$

The conditional covariance matrix $C_\sigma = \langle \mathbf{x} \cdot \mathbf{x}^T \rangle_\sigma$ can be written in terms of order parameters as

$$C_\sigma = v_\sigma \begin{pmatrix} Q_{11} & \cdots & Q_{1K} & R_{11} & \cdots & R_{1M} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ Q_{1K} & \cdots & Q_{KK} & R_{K1} & \cdots & R_{KM} \\ R_{11} & \cdots & R_{K1} & T_{11} & \cdots & T_{1M} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ R_{1M} & \cdots & R_{KM} & T_{M1} & \cdots & T_{MM} \end{pmatrix} \quad (3.29)$$

3.B Form of the Differential Equations

In order to perform the ordinary differential equations described in (3.14), we need to plug in the values of

$$\langle f_S \rangle, \quad \langle \mathbf{x}_n f_S \rangle \quad \text{and} \quad \langle f_S f_T \rangle \quad (3.30)$$

Note that $\langle f_S f_T \rangle$ is not required in the limit $\eta \to 0$, where terms proportional to η^2 can be neglected. We write the forms for the following algorithms: LVQ 2.1, LFM-W, GLVQ and RSLVQ.

LVQ 2.1

The general modulation function for LVQ 2.1 is described in Equation (3.4) as

$$f_S = \chi(c_S, y_\sigma^\mu) \sum_{T: c_T \neq c_S} \left(\Theta_{ST}^{-\delta} - \Theta_{ST}^{+\delta} \right) \prod_{U \neq S,T} \Theta_{SU} \Theta_{TU},$$

with $\chi(c_S, y_\sigma^\mu) = 1$ if $c_S = y_\sigma^\mu$ and $\chi(c_S, y_\sigma^\mu) = -1$ else. We can rewrite

$$\begin{aligned} \Theta_{ST}^{\delta} &= \Theta\left(d_T - d_S - \delta\right) \\ &= \Theta\left(-2\mathbf{w}_T \cdot \xi^\mu + \mathbf{w}_T^2 + 2\mathbf{w}_S \cdot \xi^\mu - \mathbf{w}_S^2 - \delta\right) \\ &= \Theta\left(-2h_T^\mu + 2h_S^\mu + Q_{TT} - Q_{SS} - \delta\right) \\ &= \Theta\left(\alpha_{ST} \cdot \mathbf{x} - \beta_{ST}^\delta\right), \end{aligned} \quad (3.31)$$

with $\alpha_{ST} = (0, \ldots, \underbrace{+2}_{\text{at } S}, \ldots, \underbrace{-2}_{\text{at } T}, \ldots, 0)$ and $\beta_{ST}^\delta = Q_{SS} - Q_{TT} - \delta$.

For two prototype systems with labels $\mathbf{w}_S$ and $\mathbf{w}_T$, we can simplify the above as

$$f_S \quad = \quad \chi(c_S, y_\sigma^\mu)\left(\Theta_{ST}^{-\delta} - \Theta_{ST}^{+\delta}\right). \tag{3.32}$$

And the required averages over the joint density (3.30) are calculated as

$$\begin{aligned}
\langle f_S \rangle &= \left\langle \chi(c_S, y_\sigma^\mu)\left(\Theta_{ST}^{-\delta} - \Theta_{ST}^{+\delta}\right)\right\rangle = \sum_{\sigma=1}^{M} p_\sigma \chi(c_S, y_\sigma)\left\langle \Theta_{ST}^{-\delta} - \Theta_{ST}^{+\delta}\right\rangle_\sigma \\
\langle \mathbf{x}_n f_S \rangle &= \sum_{\sigma=1}^{M} p_\sigma \chi(c_S, y_\sigma)\left\langle \mathbf{x}_n(\Theta_{ST}^{-\delta} - \Theta_{ST}^{+\delta})\right\rangle_\sigma \\
\langle f_S f_S \rangle &= \left\langle \chi(c_S, y_\sigma^\mu)^2\left(\Theta_{ST}^{-\delta} - \Theta_{ST}^{+\delta}\right)^2\right\rangle = \sum_{\sigma=1}^{M} p_\sigma \left\langle \Theta_{ST}^{-\delta} - \Theta_{ST}^{+\delta}\right\rangle_\sigma \\
\langle f_S f_T \rangle &= \left\langle \chi(c_S, y_\sigma^\mu)\chi(c_T, y_\sigma^\mu)\left(\Theta_{ST}^{-\delta} - \Theta_{ST}^{+\delta}\right)^2\right\rangle = -\sum_{\sigma=1}^{M} p_\sigma \left\langle \Theta_{ST}^{-\delta} - \Theta_{ST}^{+\delta}\right\rangle_\sigma
\end{aligned} \tag{3.33}$$

The quantities $\langle(\Theta_{ST}^{-\delta} - \Theta_{ST}^{+\delta})\rangle_\sigma$ and $\langle \mathbf{x}_n(\Theta_{ST}^{-\delta} - \Theta_{ST}^{+\delta})\rangle_\sigma$ are calculated in Appendix 3.C.

LFM-W

The general modulation function for LFM-W is described in Equation (3.5) as

$$f_S \quad = \quad \begin{cases} \sum\limits_{K:c_K \neq y_\sigma} \left(\Theta_{KS} - \Theta_{KS}^{\delta}\right)\psi(S,K) & \text{if } c_S = y_\sigma^\mu \\ \sum\limits_{J:c_J = y_\sigma} \left(\Theta_{SJ} - \Theta_{SJ}^{\delta}\right)\psi(J,S) & \text{else.} \end{cases} \tag{3.34}$$

with $\psi(J,K) = \prod_{T:c_T = y_\sigma} \Theta_{JT} \prod_{U:c_U \neq y_\sigma} \Theta_{KU}$. With only two prototypes, both $\mathbf{w}_S$ and $\mathbf{w}_T$ are winners of their respective class, thus $\psi(.) = 1$ and the averages are

$$\begin{aligned}
\langle f_S \rangle &= \sum_{\sigma:y_\sigma = c_S}^{M} p_\sigma \langle \Theta_{TS} - \Theta_{TS}^{\delta}\rangle_\sigma + \sum_{\sigma:y_\sigma \neq c_S}^{M} p_\sigma \langle \Theta_{ST} - \Theta_{ST}^{\delta}\rangle_\sigma \\
\langle \mathbf{x}_n f_S \rangle &= \sum_{\sigma:y_\sigma = c_S}^{M} p_\sigma \langle \mathbf{x}_n(\Theta_{TS} - \Theta_{TS}^{\delta})\rangle_\sigma + \sum_{\sigma:y_\sigma \neq c_S}^{M} p_\sigma \langle \mathbf{x}_n(\Theta_{ST} - \Theta_{ST}^{\delta})\rangle_\sigma
\end{aligned} \tag{3.35}$$

GLVQ

The general modulation function for GLVQ is described in Equation (3.9) as

$$f_S = \begin{cases} \sum_{K:c_K \neq y_\sigma} \left(\frac{2}{v_G} \phi \left(\frac{d_S - d_K}{v_G} \right) \right) \psi(S,K) & \text{if } c_S = y_\sigma^\mu \\ -\sum_{J:c_J = y_\sigma} \left(\frac{2}{v_G} \phi \left(\frac{d_J - d_S}{v_G} \right) \right) \psi(J,S) & \text{else.} \end{cases} \tag{3.36}$$

For two prototypes,

$$\begin{aligned} \langle f_S \rangle &= \sum_{\sigma:y_\sigma = c_S}^{M} p_\sigma \frac{2}{v_G} \langle \phi(\alpha_{ST} \cdot \mathbf{x} - \beta_{ST}) \rangle_\sigma - \sum_{\sigma:y_\sigma \neq c_S}^{M} p_\sigma \frac{2}{v_G} \langle \phi(\alpha_{ST} \cdot \mathbf{x} - \beta_{ST}) \rangle_\sigma \\ \langle \mathbf{x}_n f_S \rangle &= \sum_{\sigma:y_\sigma = c_S}^{M} p_\sigma \frac{2}{v_G} \langle \phi(\alpha_{ST} \cdot \mathbf{x} - \beta_{ST}) \rangle_\sigma - \sum_{\sigma:y_\sigma \neq c_S}^{M} p_\sigma \frac{2}{v_G} \langle \phi(\alpha_{ST} \cdot \mathbf{x} - \beta_{ST}) \rangle_\sigma \end{aligned} \tag{3.37}$$

$$\text{with } \alpha_{ST} = \{\ldots, \underbrace{-\frac{2}{v_G}}_{\text{at } S}, \ldots, \underbrace{+\frac{2}{v_G}}_{\text{at } T}, \ldots 0, 0\}, \quad \beta_{ST} = -\frac{Q_{SS} - Q_{TT}}{v_G}.$$

The quantities $\langle \phi(\alpha_{ST} \cdot \mathbf{x} - \beta_{ST}) \rangle_\sigma$ are found in Equation (3.48) in Appendix 3.C.

RSLVQ

With one prototype representing each class, (2.12) become

$$\begin{aligned} P_\sigma(S|\xi^\mu) &= \frac{\exp\left(-(\xi^\mu - \mathbf{w}_S^\mu)^2/2v_{\text{soft}}\right)}{\exp\left(-(\xi^\mu - \mathbf{w}_S^\mu)^2/2v_{\text{soft}}\right)} = 1 \\ P(S|\xi^\mu) &= \frac{\exp\left(-(\xi^\mu - \mathbf{w}_S^\mu)^2/2v_{\text{soft}}\right)}{\sum_{T=1}^{K} \exp\left(-(\xi^\mu - \mathbf{w}_T^\mu)^2/2v_{\text{soft}}\right)} \\ &= \frac{1}{1 + \sum_{T \neq S}^{K} \exp\left(\frac{1}{2v_{\text{soft}}}\left(-2\xi^\mu \mathbf{w}_S^\mu + (\mathbf{w}_S^\mu)^2 + 2\xi^\mu \mathbf{w}_T^\mu - (\mathbf{w}_T^\mu)^2\right)\right)} \\ &= \frac{1}{1 + \sum_{T \neq S}^{K} \exp\left(\frac{1}{2v_{\text{soft}}}\left(-2h_S + Q_{SS} + 2h_T - Q_{TT}\right)\right)} \\ &= \frac{1}{1 + \sum_{T \neq S}^{K} \exp\left(\alpha_{ST} \cdot \mathbf{x} - \beta_{ST}\right)} \end{aligned} \tag{3.38}$$

where we defined

$$\alpha_{ST} = \{\ldots, \underbrace{-\frac{1}{v_{\text{soft}}}}_{\text{at } S}, \ldots, \underbrace{+\frac{1}{v_{\text{soft}}}}_{\text{at } T}, \ldots 0, 0\}, \quad \beta_{ST} = -\frac{Q_{SS} - Q_{TT}}{2v_{\text{soft}}}$$

Therefore the RSLVQ modulation function becomes

$$f_S = \frac{1}{v_{\text{soft}}} \left(\delta(c_S, y_\sigma^\mu) - \Omega_S\right) \tag{3.39}$$

where $\delta(x, y)$ is the Kronecker delta and

$$\Omega_S = \frac{1}{1 + \sum_{T \neq S}^{K} \exp\left(\alpha_{ST} \cdot \mathbf{x} - \beta_{ST}\right)} \tag{3.40}$$

We obtain the averages

$$\begin{aligned}
\langle f_S \rangle &= \frac{1}{v_{\text{soft}}} \langle \delta(c_S, y_\sigma) - \Omega_S \rangle = \frac{1}{v_{\text{soft}}} \left(\sum_{\sigma: y_\sigma = c_S} p_\sigma - \sum_\sigma p_\sigma \langle \Omega_S \rangle_\sigma \right) \\
\langle \mathbf{x}_n f_S \rangle &= \begin{cases} \frac{1}{v_{\text{soft}}} \left(\sum_{\sigma: y_\sigma = c_S} p_\sigma \langle h_n \rangle_\sigma - \sum_\sigma p_\sigma \langle \mathbf{x}_n \Omega_S \rangle_\sigma \right) & \text{if } n \leq K \\ \frac{1}{v_{\text{soft}}} \left(\sum_{\sigma: y_\sigma = c_S} p_\sigma \langle b_{n-K} \rangle_\sigma - \sum_\sigma p_\sigma \langle \mathbf{x}_n \Omega_S \rangle_\sigma \right) & \text{if } n > K \end{cases} \\
&= \begin{cases} \frac{1}{v_{\text{soft}}} \left(\sum_{\sigma: y_\sigma = c_S} p_\sigma \ell_\sigma R_{n\sigma} - \sum_\sigma p_\sigma \langle \mathbf{x}_n \Omega_S \rangle_\sigma \right) & \text{if } n \leq K \\ \frac{1}{v_{\text{soft}}} \left(\sum_{\sigma: y_\sigma = c_S} p_\sigma \ell_\sigma T_{(n-K)\sigma} - \sum_\sigma p_\sigma \langle \mathbf{x}_n \Omega_S \rangle_\sigma \right) & \text{if } n > K \end{cases}
\end{aligned} \tag{3.41}$$

The required quantities $\langle \Omega_S \rangle_\sigma$ and $\langle \mathbf{x}_n \Omega_S \rangle_\sigma$ are supplied in Appendix 3.C.

3.C Gaussian Averages

3.C.1 Two prototypes

For generic functions $f_{ab} \equiv f(\alpha_{ab} \cdot \mathbf{x} - \beta_{ab})$, the quantities $\langle f_{ab} \rangle_\sigma$ and $\langle \mathbf{x}_n f_{ab} \rangle_\sigma$ are required.

$$\begin{aligned}
\langle f_{ab} \rangle_\sigma &= \frac{1}{(2\pi)^{D/2}(\det(C_\sigma))^{1/2}} \int_{\mathbb{R}^D} f(\alpha_{ab} \cdot \mathbf{x} - \beta_{ab}) \\
&\quad \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T C_\sigma^{-1}(\mathbf{x}-\mu)\right) d\mathbf{x} \\
&= \frac{1}{(2\pi)^{D/2}(\det(C_\sigma))^{1/2}} \int_{\mathbb{R}^D} f(\alpha_{ab} \cdot \mathbf{x}' + \alpha_{ab} \cdot \mu - \beta_{ab}) \\
&\quad \exp\left(-\frac{1}{2}(\mathbf{x}')^T C_\sigma^{-1}(\mathbf{x}')\right) d\mathbf{x}' \\
&= \frac{1}{(2\pi)^{D/2}} \int_{\mathbb{R}^D} f\left(\alpha_{ab} C_\sigma^{-1/2}\mathbf{y} + \widetilde{\beta}_{ab,\sigma}\right) \exp\left(-\frac{1}{2}\mathbf{y}^2\right) d\mathbf{y} \qquad (3.42)
\end{aligned}$$

with $\widetilde{\beta}_{ab,\sigma} = \alpha_{ab} \cdot \mathbf{x} - \beta_{ab}$. Rotating the coordinate system, we obtain

$$\begin{aligned}
\langle f_{ab} \rangle_\sigma &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f\left(\|\alpha_{ab} C_\sigma^{-1/2}\| \widetilde{y} + \widetilde{\beta}_{ab,\sigma}\right) \exp\left(-\frac{1}{2}\widetilde{y}^2\right) d\widetilde{y} \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f\left(\widetilde{\alpha}_{ab,\sigma}\widetilde{y} + \widetilde{\beta}_{ab,\sigma}\right) \exp\left(-\frac{1}{2}\widetilde{y}^2\right) d\widetilde{y} \qquad (3.43)
\end{aligned}$$

with $\widetilde{\alpha}_{ab,\sigma} = \|\alpha_{ab} C_\sigma^{-1/2}\|$. Next we calculate the quantity

$$\begin{aligned}
\langle \mathbf{x}_n f_{ab} \rangle_\sigma &= \frac{1}{(2\pi)^{D/2}(\det(C_\sigma))^{1/2}} \int_{\mathbb{R}^D} \mathbf{x}_n f(\alpha_{ab} \cdot \mathbf{x} - \beta_{ab}) \\
&\quad \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T C_\sigma^{-1}(\mathbf{x}-\mu)\right) d\mathbf{x} \\
&= \frac{1}{(2\pi)^{D/2}} \int_{\mathbb{R}^D} \left(C_\sigma^{1/2}\mathbf{y} + \mu\right)_n f\left(\alpha_{ab} C_\sigma^{-1/2}\mathbf{y} + \widetilde{\beta}_{ab,\sigma}\right) \exp\left(-\frac{1}{2}\mathbf{y}^2\right) d\mathbf{y} \\
&= \frac{1}{(2\pi)^{D/2}} \int_{\mathbb{R}^D} \left(C_\sigma^{1/2}\mathbf{y}\right)_n f\left(\alpha_{ab} C_\sigma^{-1/2}\mathbf{y} + \widetilde{\beta}_{ab,\sigma}\right) \exp\left(-\frac{1}{2}\mathbf{y}^2\right) d\mathbf{y} \\
&\quad + (\mu)_n \langle f_{ab} \rangle_\sigma \qquad (3.44)
\end{aligned}$$

LVQ 2.1, LFM-W

The following quantities are required for two prototype LVQ 2.1 and LFM-W:

$$\left\langle \Theta_{ab}^\delta - \Theta_{ab}^\gamma \right\rangle_\sigma = \Phi\left(\frac{\widetilde{\beta}_{ab,\sigma}^\delta}{\widetilde{\alpha}_{ab,\sigma}}\right) - \Phi\left(\frac{\widetilde{\beta}_{ab,\sigma}^\gamma}{\widetilde{\alpha}_{ab,\sigma}}\right) \qquad (3.45)$$

$$\begin{aligned}\left\langle \mathbf{x}_n \left(\Theta_{ab}^{\delta} - \Theta_{ab}^{\gamma}\right)\right\rangle_{\sigma} &= \frac{(C_\sigma \alpha_{ab})_n}{\sqrt{2\pi}\widetilde{\alpha}_{ab,\sigma}} \left\{ \exp\left[-\frac{1}{2}\left(\frac{\widetilde{\beta}_{ab,\sigma}^{\delta}}{\widetilde{\alpha}_{ab,\sigma}}\right)^2\right] - \exp\left[-\frac{1}{2}\left(\frac{\widetilde{\beta}_{ab,\sigma}^{\gamma}}{\widetilde{\alpha}_{ab,\sigma}}\right)^2\right]\right\} \\ &\quad + (\mu_\sigma)_n \left[\Phi\left(\frac{\widetilde{\beta}_{ab,\sigma}^{\delta}}{\widetilde{\alpha}_{ab,\sigma}}\right) - \Phi\left(\frac{\widetilde{\beta}_{ab,\sigma}^{\gamma}}{\widetilde{\alpha}_{ab,\sigma}}\right)\right] \end{aligned} \tag{3.46}$$

GLVQ

For GLVQ, the quantities $\langle \phi_{ab} \rangle_\sigma$ and $\langle \mathbf{x}_n \phi_{ab} \rangle_\sigma$ are required.

$$\begin{aligned}\langle \phi_{ab} \rangle_\sigma &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \phi\left(\widetilde{\alpha}_{ab,\sigma}\widetilde{y} + \widetilde{\beta}_{ab,\sigma}\right) \exp\left(-\frac{1}{2}\widetilde{y}^2\right) d\widetilde{y} \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\widetilde{\beta}_{ab,\sigma}^2\right) \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\widetilde{\alpha}_{ab,\sigma}^2 + 1)\widetilde{y}^2 - \widetilde{\alpha}_{ab,\sigma}\widetilde{\beta}_{ab,\sigma}\widetilde{y}\right) d\widetilde{y} \end{aligned} \tag{3.47}$$

Here we can use the substitution $\int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}ax^2 + bx\right) = \frac{1}{\sqrt{a}} \exp\left(\frac{b^2}{2a}\right)$ to obtain

$$\begin{aligned}\langle \phi_{ab} \rangle_\sigma &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\widetilde{\beta}_{ab,\sigma}^2\right) \frac{1}{\sqrt{\widetilde{\alpha}_{ab,\sigma}^2 + 1}} \exp\left(\frac{\widetilde{\alpha}_{ab,\sigma}^2 \widetilde{\beta}_{ab,\sigma}^2}{2(\widetilde{\alpha}_{ab,\sigma}^2 + 1)}\right) \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\widetilde{\alpha}_{ab,\sigma}^2 + 1}} \exp\left(-\frac{1}{2}\widetilde{\beta}_{ab,\sigma}^2 \left(1 - \frac{\widetilde{\alpha}_{ab,\sigma}^2}{(\widetilde{\alpha}_{ab,\sigma}^2 + 1)}\right)\right) \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\widetilde{\alpha}_{ab,\sigma}^2 + 1}} \exp\left(-\frac{1}{2}\frac{\widetilde{\beta}_{ab,\sigma}^2}{\widetilde{\alpha}_{ab,\sigma}^2 + 1}\right) \end{aligned} \tag{3.48}$$

RSLVQ

For RSLVQ, the quantities $\langle \Omega_{ab} \rangle_\sigma$ and $\langle \mathbf{x}_n \Omega_{ab} \rangle_\sigma$ are required.

$$\langle \Omega_{ab} \rangle_\sigma = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{1}{1 + \exp(\widetilde{\alpha}_{ab,\sigma}\widetilde{y} + \widetilde{\beta}_{ab,\sigma})} \exp\left(-\frac{1}{2}\widetilde{y}^2\right) d\widetilde{y} \tag{3.49}$$

This one-dim. integration has to be solved numerically.

$$\begin{aligned}\langle \mathbf{x}_n \Omega_{ab} \rangle_\sigma &= \frac{1}{(2\pi)^{D/2}} \int_{\mathbb{R}^D} \frac{\left(C_\sigma^{1/2}\mathbf{y}\right)_n}{1+\exp(\alpha_{ab} C_\sigma^{-1/2}\mathbf{y} + \widetilde{\beta}_{ab,\sigma})} \exp\left(-\frac{1}{2}\mathbf{y}^2\right) d\mathbf{y} \\ &\quad + (\mu)_n \langle \Omega_{ab} \rangle_\sigma \\ &= \frac{1}{(2\pi)^{D/2}} \sum_{j=1}^{D} I_j + (\mu)_n \langle \Omega_{ab} \rangle_\sigma \end{aligned} \tag{3.50}$$

where

$$I_j = \frac{1}{(2\pi)^{D/2}} \int_{\mathbb{R}} \frac{(C_\sigma^{1/2})_{nj}(\mathbf{y})_n}{1+\exp(\alpha_{ab} C_\sigma^{-1/2}\mathbf{y} + \widetilde{\beta}_{ab,\sigma})} \exp\left(-\frac{1}{2}(\mathbf{y}_j)^2\right) d(\mathbf{y})_j \tag{3.51}$$

Applying integration by parts $\int u dv = uv - \int v du$ with

$$\begin{aligned} u &= \frac{1}{1+\exp(\alpha_{ab} C_\sigma^{-1/2}\mathbf{y} + \widetilde{\beta}_{ab,\sigma})} \\ v &= (C_\sigma^{1/2})_{nj} \exp\left(-\frac{1}{2}(\mathbf{y}_j)^2\right) \\ du &= -\frac{\exp(\alpha_{ab} C_\sigma^{-1/2}\mathbf{y} + \widetilde{\beta}_{ab,\sigma})}{\left(1+\exp(\alpha_{ab} C_\sigma^{-1/2}\mathbf{y} + \widetilde{\beta}_{ab,\sigma})\right)^2} \frac{\partial}{\partial(\mathbf{y})_j} \left(\alpha_{ab} C_\sigma^{-1/2}\mathbf{y}\right) d(\mathbf{y})_j \\ dv &= -(C_\sigma^{1/2})_{nj}(\mathbf{y})_j \exp\left(-\frac{1}{2}(\mathbf{y}_j)^2\right) d(\mathbf{y})_j, \end{aligned} \tag{3.52}$$

we obtain

$$\begin{aligned} I_j &= \left[\frac{1}{1+\exp(\alpha_{ab} C_\sigma^{-1/2}\mathbf{y} + \widetilde{\beta}_{ab,\sigma})} (C_\sigma^{1/2})_{nj} \exp\left(-\frac{1}{2}(\mathbf{y}_j)^2\right)\right]_{-\infty}^{\infty} \\ &\quad - \int_{\mathbb{R}} \frac{(C_\sigma^{1/2})_{nj} \exp(\alpha_{ab} C_\sigma^{-1/2}\mathbf{y} + \widetilde{\beta}_{ab,\sigma})}{\left(1+\exp(\alpha_{ab} C_\sigma^{-1/2}\mathbf{y} + \widetilde{\beta}_{ab,\sigma})\right)^2} \frac{\partial}{\partial(\mathbf{y})_j} \left(\alpha_{ab} C_\sigma^{-1/2}\mathbf{y}\right) \exp\left(-\frac{1}{2}(\mathbf{y}_j)^2\right) d(\mathbf{y})_j \end{aligned} \tag{3.53}$$

$$\begin{aligned}\langle \mathbf{x}_n \Omega_{ab} \rangle_\sigma &= -\frac{1}{(2\pi)^{D/2}} \sum_{j=1}^{D} \int_{\mathbb{R}} \frac{(C_\sigma^{1/2})_{nj} \exp\left(\alpha_{ab} C_\sigma^{-1/2} \mathbf{y} + \widetilde{\beta}_{ab,\sigma}\right)}{\left(1 + \exp\left(\alpha_{ab} C_\sigma^{-1/2} \mathbf{y} + \widetilde{\beta}_{ab,\sigma}\right)\right)^2} \\ &\quad \frac{\partial}{\partial(\mathbf{y})_j}\left(\alpha_{ab} C_\sigma^{-1/2} \mathbf{y}\right) \exp\left(-\frac{1}{2}(\mathbf{y}_j)^2\right) d(\mathbf{y})_j + (\mu)_n \langle \Omega_{ab} \rangle_\sigma \\ &= -\frac{1}{(2\pi)^{D/2}} (C_k \alpha_{ab})_n \int_{\mathbb{R}} \frac{\exp\left(\alpha_{ab} C_\sigma^{-1/2} \mathbf{y} + \widetilde{\beta}_{ab,\sigma}\right)}{\left(1 + \exp\left((\alpha_{ab} C_\sigma^{-1/2} \mathbf{y} + \widetilde{\beta}_{ab,\sigma}\right)\right)^2} \\ &\quad \exp\left(-\frac{1}{2}(\mathbf{y}_j)^2\right) d(\mathbf{y})_j \end{aligned} \tag{3.54}$$

After applying rotation,

$$\begin{aligned}\langle \mathbf{x}_n \Omega_{ab} \rangle_\sigma &= -\frac{(C_k \alpha_{ab})_n}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{\exp \|\alpha_{ab} C_\sigma^{-1/2}\| \widetilde{y} + \widetilde{\beta}_{ab,\sigma})}{\left(1 + \exp\left(\|\alpha_{ab} C_\sigma^{-1/2}\| \widetilde{y} + \widetilde{\beta}_{ab,\sigma}\right)\right)^2} \\ &\quad \times \exp\left(-\frac{1}{2}\widetilde{y}^2\right) d\widetilde{y} + (\mu_k)_n \langle \Omega_{ab} \rangle_\sigma \\ &= -\frac{(C_k \alpha_{ab})_n}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{\exp\left(\widetilde{\alpha}_{ab,\sigma} \widetilde{y} + \widetilde{\beta}_{ab,\sigma}\right)}{\left(1 + \exp\left(\widetilde{\alpha}_{ab,\sigma} \widetilde{y} + \widetilde{\beta}_{ab,\sigma}\right)\right)^2} \exp\left(-\frac{1}{2}\widetilde{y}^2\right) d\widetilde{y} \\ &\quad + (\mu_k)_n \langle \Omega_{ab} \rangle_\sigma \end{aligned} \tag{3.55}$$

which is also solved numerically.

3.C.2 Three prototypes

For generic function $f_{ab} f_{cd} \equiv f(\alpha_{ab} \cdot \mathbf{x} - \beta_{ab}) f(\alpha_{cd} \cdot \mathbf{x} - \beta_{cd})$, the quantities $\langle f_{ab} f_{cd} \rangle_k$ and $\langle \mathbf{x}_n f_{ab} f_{cd} \rangle_k$ are required.

$$\begin{aligned}\langle f_{ab} f_{cd} \rangle_\sigma &= \frac{1}{(2\pi)^{D/2}} \int_{\mathbb{R}^D} f\left(\alpha_{ab} C_\sigma^{-1/2} \mathbf{y} + \widetilde{\beta}_{ab,\sigma}\right) f\left(\alpha_{cd} C_\sigma^{-1/2} \mathbf{y} + \widetilde{\beta}_{cd}\right) \\ &\quad \times \exp\left(-\frac{1}{2}\mathbf{y}^2\right) d\mathbf{y} \end{aligned} \tag{3.56}$$

Next we calculate the quantity

$$\begin{aligned}\langle \mathbf{x}_n f_{ab} f_{cd} \rangle_\sigma &= \frac{1}{(2\pi)^{D/2}} \int_{\mathbb{R}^D} \left(C_\sigma^{1/2} \mathbf{y}\right)_n f\left(\alpha_{ab} C_\sigma^{-1/2} \mathbf{y} + \widetilde{\beta}_{ab,\sigma}\right) f\left(\alpha_{cd} C_\sigma^{-1/2} \mathbf{y} + \widetilde{\beta}_{cd}\right) \\ &\quad \times \exp\left(-\frac{1}{2}\mathbf{y}^2\right) d\mathbf{y} + (\mu)_n \langle f_{ab} f_{cd} \rangle_\sigma \end{aligned} \tag{3.57}$$

The quantities $\langle \Theta_{ab}\Theta_{cd}\rangle_\sigma$ and $\langle \mathbf{x}_n\Theta_{ab}\Theta_{cd}\rangle_\sigma$ have been calculated in Witoelar et al. (2008), as follows:

$$\langle \Theta_{ab}\Theta_{cd}\rangle_\sigma = \frac{1}{\sqrt{2\pi}} \int_{-\frac{\widetilde{\beta}_{ab,\sigma}}{\widetilde{\alpha}_{ab,\sigma}}}^{\infty} \exp\left(-\frac{1}{2}y_1'^2\right) \Phi\left(\frac{\widetilde{\beta}_{cd,\sigma}\widetilde{\alpha}_{ab,\sigma} + (\alpha_{cd}C_\sigma\alpha_{ab})y_1'}{\sqrt{\widetilde{\alpha}_{cd,\sigma}^2\widetilde{\alpha}_{ab,\sigma}^2 - (\alpha_{cd}C_\sigma\alpha_{ab})^2}}\right) dy_1' \tag{3.58}$$

$$\begin{aligned}\langle (\mathbf{x})_n\Theta_{ab}\Theta_{cd}\rangle_\sigma &= \frac{(C_\sigma\alpha_{ab})_n}{\sqrt{(2\pi)}\widetilde{\alpha}_{ab,\sigma}} \exp\left(-\frac{1}{2}\frac{\widetilde{\beta}_{ab,\sigma}^2}{\widetilde{\alpha}_{ab,\sigma}^2}\right) \Phi\left(\frac{\widetilde{\beta}_{cd,\sigma}\widetilde{\alpha}_{ab,\sigma}^2 - \widetilde{\beta}_{ab,\sigma}(\alpha_{cd}C_\sigma\alpha_{ab})}{\widetilde{\alpha}_{ab,\sigma}\sqrt{\widetilde{\alpha}_{cd,\sigma}^2\widetilde{\alpha}_{ab,\sigma}^2 - (\alpha_{cd}C_\sigma\alpha_{ab})^2}}\right) \\ &+ \frac{(C_\sigma\alpha_{cd})_n}{\sqrt{(2\pi)}\widetilde{\alpha}_{cd,\sigma}} \exp\left(-\frac{1}{2}\frac{\widetilde{\beta}_{cd,\sigma}^2}{\widetilde{\alpha}_{cd,\sigma}^2}\right) \Phi\left(\frac{\widetilde{\beta}_{ab,\sigma}\widetilde{\alpha}_{cd,\sigma}^2 - \widetilde{\beta}_{cd,\sigma}(\alpha_{ab}C_\sigma\alpha_{cd})}{\widetilde{\alpha}_{cd,\sigma}\sqrt{\widetilde{\alpha}_{cd,\sigma}^2\widetilde{\alpha}_{ab,\sigma}^2 - (\alpha_{cd}C_\sigma\alpha_{ab})^2}}\right) \\ &+ (\mu_\sigma)_n\langle \Theta_{ab}\Theta_{cd}\rangle_\sigma. \end{aligned} \tag{3.59}$$

With the addition of a window, these quantities are required:

$$\langle (\Theta_{ab}^\delta - \Theta_{ab}^\gamma)\Theta_{cd}\rangle_\sigma = \frac{1}{\sqrt{2\pi}} \int_{-\frac{\widetilde{\beta}_{ab,\sigma}^\delta}{\widetilde{\alpha}_{ab,\sigma}}}^{-\frac{\widetilde{\beta}_{ab,\sigma}^\gamma}{\widetilde{\alpha}_{ab,\sigma}}} \exp\left(-\frac{1}{2}y_1'^2\right) \Phi\left(\frac{\widetilde{\beta}_{cd,\sigma}\widetilde{\alpha}_{ab,\sigma} + (\alpha_{cd}C_\sigma\alpha_{ab})y_1'}{\sqrt{\widetilde{\alpha}_{cd,\sigma}^2\widetilde{\alpha}_{ab,\sigma}^2 - (\alpha_{cd}C_\sigma\alpha_{ab})^2}}\right) dy_1'$$

$$\begin{aligned}&\langle (\mathbf{x})_n(\Theta_{ab}^\delta - \Theta_{ab}^\gamma)\Theta_{cd}\rangle_\sigma \\ &= \frac{(C_\sigma\alpha_{ab})_n}{\sqrt{(2\pi)}\widetilde{\alpha}_{ab,\sigma}} \exp\left(-\frac{1}{2}\frac{(\widetilde{\beta}_{ab,\sigma}^\delta)^2}{\widetilde{\alpha}_{ab,\sigma}^2}\right) \Phi\left(\frac{\widetilde{\beta}_{cd,\sigma}\widetilde{\alpha}_{ab,\sigma}^2 - \widetilde{\beta}_{ab,\sigma}^\delta(\alpha_{cd}C_\sigma\alpha_{ab})}{\widetilde{\alpha}_{ab,\sigma}\sqrt{\widetilde{\alpha}_{cd,\sigma}^2\widetilde{\alpha}_{ab,\sigma}^2 - (\alpha_{cd}C_\sigma\alpha_{ab})^2}}\right) \\ &+ \frac{(C_\sigma\alpha_{cd})_n}{\sqrt{(2\pi)}\widetilde{\alpha}_{cd,\sigma}} \exp\left(-\frac{1}{2}\frac{\widetilde{\beta}_{cd,\sigma}^2}{\widetilde{\alpha}_{cd,\sigma}^2}\right) \Phi\left(\frac{\widetilde{\beta}_{ab,\sigma}^\delta\widetilde{\alpha}_{cd,\sigma}^2 - \widetilde{\beta}_{cd,\sigma}(\alpha_{ab}C_\sigma\alpha_{cd})}{\widetilde{\alpha}_{cd,\sigma}\sqrt{\widetilde{\alpha}_{cd,\sigma}^2\widetilde{\alpha}_{ab,\sigma}^2 - (\alpha_{cd}C_\sigma\alpha_{ab})^2}}\right) \\ &- \frac{(C_\sigma\alpha_{ab})_n}{\sqrt{(2\pi)}\widetilde{\alpha}_{ab,\sigma}} \exp\left(-\frac{1}{2}\frac{(\widetilde{\beta}_{ab,\sigma}^\gamma)^2}{\widetilde{\alpha}_{ab,\sigma}^2}\right) \Phi\left(\frac{\widetilde{\beta}_{cd,\sigma}\widetilde{\alpha}_{ab,\sigma}^2 - \widetilde{\beta}_{ab,\sigma}^\gamma(\alpha_{cd}C_\sigma\alpha_{ab})}{\widetilde{\alpha}_{ab,\sigma}\sqrt{\widetilde{\alpha}_{cd,\sigma}^2\widetilde{\alpha}_{ab,\sigma}^2 - (\alpha_{cd}C_\sigma\alpha_{ab})^2}}\right) \\ &- \frac{(C_\sigma\alpha_{cd})_n}{\sqrt{(2\pi)}\widetilde{\alpha}_{cd,\sigma}} \exp\left(-\frac{1}{2}\frac{\widetilde{\beta}_{cd,\sigma}^2}{\widetilde{\alpha}_{cd,\sigma}^2}\right) \Phi\left(\frac{\widetilde{\beta}_{ab,\sigma}^\gamma\widetilde{\alpha}_{cd,\sigma}^2 - \widetilde{\beta}_{cd,\sigma}(\alpha_{ab}C_\sigma\alpha_{cd})}{\widetilde{\alpha}_{cd,\sigma}\sqrt{\widetilde{\alpha}_{cd,\sigma}^2\widetilde{\alpha}_{ab,\sigma}^2 - (\alpha_{cd}C_\sigma\alpha_{ab})^2}}\right) \\ &+ (\mu_\sigma)_n\langle (\Theta_{ab}^\delta - \Theta_{ab}^\gamma)\Theta_{cd}\rangle_\sigma. \end{aligned} \tag{3.60}$$

For LVQ 2.1, the following average is required:

$$\langle (\Theta_{ab}^\delta - \Theta_{ab}^\gamma)\Theta_{ac}\Theta_{bc}\rangle_\sigma = \frac{1}{\sqrt{2\pi}} \int_{y_{1,\min}}^{y_{1,\max}} \exp\left(-\frac{1}{2}y_1'^2\right) \Phi(-y_2{}^*) dy_1 \tag{3.61}$$

$$\text{with} \qquad y_{1,\min} = -\frac{\widetilde{\beta}^{\delta}_{ab,\sigma}}{\widetilde{\alpha}_{ab,\sigma}}, \quad y_{1,\max} = -\frac{\widetilde{\beta}^{\gamma}_{ab,\sigma}}{\widetilde{\alpha}_{ab,\sigma}}$$
$$y_2{}^* = \min\left(\frac{-\widetilde{\beta}_{ac} - (\alpha_{ac}C_\sigma^{1/2}e_1)y_1}{\alpha_{ac}C_\sigma^{1/2}e_2}, \frac{-\widetilde{\beta}_{bc} - (\alpha_{bc}C_\sigma^{1/2}e_1)y_1}{\alpha_{bc}C_\sigma^{1/2}e_2}\right) \tag{3.62}$$

$$\langle \mathbf{x}_n(\Theta^{\delta}_{ab} - \Theta^{\gamma}_{ab})\Theta_{ac}\Theta_{bc}\rangle_\sigma \;=\; I_{ab} + I_{ac} + I_{bc} + (\mu)_n\langle \mathbf{x}_n(\Theta^{\delta}_{ab} - \Theta^{\gamma}_{ab})\Theta_{ac}\Theta_{bc}\rangle_\sigma \tag{3.63}$$

where

$$\begin{aligned} I_{ab} \;=\; & \frac{(C_\sigma\alpha_{ab})_n}{\sqrt{2\pi}\widetilde{\alpha}_{ab,\sigma}}\Bigg[\exp\left(-\frac{1}{2}\left(-\frac{\widetilde{\beta}^{\delta}_{ab,\sigma}}{\widetilde{\alpha}_{ab,\sigma}}\right)^2\right)\left(\Phi(-y^{\delta}_{2,\min}) - \Phi(-y^{\delta}_{2,\max})\right) \\ & - \exp\left(-\frac{1}{2}\left(-\frac{\widetilde{\beta}^{\gamma}_{ab,\sigma}}{\widetilde{\alpha}_{ab,\sigma}}\right)^2\right)\left(\Phi(-y^{\gamma}_{2,\min}) - \Phi(-y^{\gamma}_{2,\max})\right)\Bigg] \end{aligned} \tag{3.64}$$

$$\begin{aligned} I_{ac} \;=\; & \frac{(C_\sigma\alpha_{ac})_n}{\sqrt{2\pi}\widetilde{\alpha}_{ac,\sigma}}\Bigg[\exp\left(-\frac{1}{2}(z)^2\right)\left(\Phi(-y_{2,\min}) - \Phi(-y_{2,\max})\right) \\ & - \exp\left(-\frac{1}{2}(z)^2\right)\left(\Phi(-y_{2,\min}) - \Phi(-y_{2,\max})\right)\Bigg] \end{aligned} \tag{3.65}$$

$$\text{with} \qquad y_{1,\min} = -\frac{\widetilde{\beta}^{\delta}_{ab,\sigma}}{\widetilde{\alpha}_{ab,\sigma}}, \quad y_{1,\max} = -\frac{\widetilde{\beta}^{\gamma}_{ab,\sigma}}{\widetilde{\alpha}_{ab,\sigma}}$$
$$y_2{}^* = \min\left(\frac{-\widetilde{\beta}_{ac} - (\alpha_{ac}C_\sigma^{1/2}e_1)y_1}{\alpha_{ac}C_\sigma^{1/2}e_2}, \frac{-\widetilde{\beta}_{bc} - (\alpha_{bc}C_\sigma^{1/2}e_1)y_1}{\alpha_{bc}C_\sigma^{1/2}e_2}\right) \tag{3.66}$$

3.D Generalization error

Two prototypes

We compute the generalization error from Equation (3.16) as follows. For two prototypes $\mathbf{w}_+$ and $\mathbf{w}_-$, we calculate $\epsilon_g = \sum p_\sigma \epsilon_{g,\sigma}$ with

$$\epsilon_{g,\sigma} = \langle \Theta_{-\sigma\sigma}\rangle_+ = \Phi\left(\frac{\widetilde{\beta}_{-\sigma\sigma,\sigma}}{\widetilde{\alpha}_{-\sigma\sigma,\sigma}}\right) \tag{3.67}$$

with $\widetilde{\alpha}_{ST,\sigma} = \sqrt{\alpha_{ST} C_\sigma \alpha_{ST}}$ and $\widetilde{\beta}_{ST,\sigma} = \alpha_{ST}\mu_\sigma - \beta_{ST}$. We refer the calculations to Biehl et al. (2004). Plugging in the values, we obtain

$$\epsilon_{g,\sigma} = \Phi\left(\frac{Q_{\sigma\sigma} - Q_{-\sigma,-\sigma} - 2\ell_\sigma(R_{\sigma,\sigma} - R_{-\sigma,\sigma})}{2\sqrt{v_\sigma}\sqrt{Q_{\sigma\sigma} - 2Q_{\sigma,-\sigma} + Q_{-\sigma,-\sigma}}}\right) \tag{3.68}$$

By using $Z_\sigma = Q_{\sigma\sigma} - Q_{-\sigma,-\sigma} - 2\ell(R_{\sigma\sigma} - R_{-\sigma\sigma})$ and $\Delta_q = \sqrt{Q_{++} - 2Q_{+-} + Q_{--}}$, we can calculate the derivative of the generalization error with respect to the order parameters $\mathbf{O} = \{R_{++}, R_{+-}, R_{-+}, R_{--}, Q_{++}, Q_{+-}, Q_{--}\}^T$ as follows:

$$\frac{d\epsilon_{g_\sigma}}{d\mathbf{O}} = \frac{1}{\sqrt{2\pi}2\sqrt{v_\sigma}} \exp\left(-\frac{1}{2}\left[\frac{Z_\sigma}{2\sqrt{v_\sigma}\Delta_q}\right]^2\right) \frac{d}{d\mathbf{O}}\frac{Z_\sigma}{\Delta_q} \tag{3.69}$$

where we used $d\Phi(\tau)/d\tau = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}\tau^2)$. Derivations with respect to the order parameters yield

$$\frac{d}{d\mathbf{O}}\frac{Z_+}{\Delta_q} = \begin{bmatrix} -2\ell/\Delta_q \\ 0 \\ +2\ell/\Delta_q \\ 0 \\ 1/\Delta_q - Z_+/(2\Delta_q^3) \\ Z_+/\Delta_q^3 \\ -1/\Delta_q - Z_+/(2\Delta_q^3) \end{bmatrix}, \quad \frac{d}{d\mathbf{O}}\frac{Z_-}{\Delta_q} = \begin{bmatrix} 0 \\ +2\ell/\Delta_q \\ 0 \\ -2\ell/\Delta_q \\ -1/\Delta_q - Z_-/(2\Delta_q^3) \\ Z_-/\Delta_q^3 \\ 1/\Delta_q - Z_-/(2\Delta_q^3) \end{bmatrix} \tag{3.70}$$

In the special case of $p_+ = p_- = 0.5$ and $v_+ = v_- = v$, one obtains

$$\frac{d\epsilon_{g_\sigma}}{d\mathbf{O}} = \sum_\sigma \frac{d\epsilon_{g_\sigma}}{d\mathbf{O}} = \frac{1}{2\sqrt{2\pi}\sqrt{v}} \exp\left(-\frac{1}{2}\left[\frac{Z}{2\sqrt{v}\Delta_q}\right]^2\right) \begin{bmatrix} -\ell/\Delta_q \\ +\ell/\Delta_q \\ +\ell/\Delta_q \\ -\ell/\Delta_q \\ -Z/(2\Delta_q^3) \\ Z/\Delta_q^3 \\ -Z/(2\Delta_q^3) \end{bmatrix}. \tag{3.71}$$

Three prototypes

To compute the generalization error in systems with three prototypes $\mathbf{w}_S, \mathbf{w}_T, \mathbf{w}_U$, we require the quantity

$$\epsilon_{g,\sigma} = \sum_{S:c_S \neq y_\sigma}^{K} \langle \Theta_{ST}\Theta_{SU}\rangle_\sigma \ , \tag{3.72}$$

where the averages are written in Equation (3.58).

Material based on:
de Vries, J. J. G., Pauws, S. C. and Biehl, M.: in press, Insightful Stress Detection from Physiology Modalities using Learning Vector Quantization, Neurocomputation

Hoofdstuk 4
EMOTION FROM A BODILY PERSPECTIVE

Abstract

Stress in daily life can lead to severe conditions as burn-out and depression and has a major impact on society. Being able to measure mental stress reliably opens up the ability to intervene in an early stage. We performed a large-scale study in which skin conductance, respiration and electrocardiogram were measured in semi-controlled conditions. Using Learning Vector Quantization techniques, we obtained up to 88% accuracy in the classification task to separate stress from relaxation. Relevance learning was used to identify the most informative features, indicating that most information is embedded in the cardiac signals. In addition to commonly used features, we also explored various novel features, of which the very-high frequency band of the power spectrum was found to be a very relevant addition.

4.1 Introduction

The harsh reality of daily life is that it becomes increasingly stressful. While certain levels of psychological stress help us perform optimally, prolonged exposure to stressors can have severe effects on wellbeing. Chronic stress is known to contribute to the development of, among others, cardiovascular diseases (Backé et al. 2012, Kivimäki et al. 2006) and has been found to contribute to high societal costs. In the US, for example, it has been estimated that job stress costs "over $300 billion annually due to increased absenteeism, employee turnover, diminished productivity, medical, legal and insurance expenses, and workers' compensation payments" (Rosch 2001).

The fine balance between the positive effects of short term stress and the detrimental effects of chronic stress on the one hand, and an increasingly demanding society on the other hand, indicate the need for assistance in balancing workload. Various products are available to help regulate mental stress (Heber et al. 2013, Westerink et al. 2014) including various biofeedback systems. One such biofeedback method is the stimulation of alpha-frequency brain waves, i.e., alpha neurofeedback (Dempster and Vernon 2009). Alpha brain waves are related to relaxation

during wake, and stimulation of these waves are known to increase relaxation levels (Gruzelier 2002). The effects have been studied in the lab quite extensively, but only limitedly in circumstances that better reflect daily life. The application of neuro-feedback in a consumer device using the paradigm of music listening was researched (van Boxtel et al. 2012) in a double-blinded experiment with two types of control, as one aim of a comprehensive study. The effectiveness of such methods can, however, be further improved by providing them at the right moment to the right people. To that end, an objective method of measuring stress using easily and unobtrusively measurable physiological parameters is needed. This lead to the second aim of the aforementioned study: the development of such a method; which is subject of the present manuscript.

Several studies have attempted to classify stress from physiological measurements (Healey and Picard 2005, Zhai et al. 2005, Zhai and Barreto 2006, Choi and Gutierrez-Osuna 2009, Wijsman et al. 2011, Giakoumis et al. 2013) using various classification techniques. Among the more popular are Support Vector Machine (SVM) and Artificial Neural Network (ANN) (Sharma and Gedeon 2012). LVQ is a relatively novel technique that has been applied successfully to a wide range of classification challenges (Neural Networks Research Centre, Helsinki 2002), but rarely to classification of affect, and to the best of our knowledge, not yet to stress classification. The family of LVQ classification techniques use prototypes that are defined in the same mathematical space as the input data. The intuitive nature and ease of inspection give LVQ an advantage over less open-box methods such as SVM and ANN. We exploit this property of LVQ to gain new insights in the field of mental stress detection, where further understanding of the domain can help improve descriptive models (Sharma and Gedeon 2012).

In the present study, we set out to build classifiers to distinguish stress from relaxation using the three modalities of Electrocardiogram (ECG), Galvanic Skin Response (GSR), and Respiration (RSP). To that end we employ LVQ methods as well as SVM. We will use these methods to explore performance of uni-modal and multimodal classifiers in order to find out which signal is most rich in information to distinguish stressful reactions and investigate how individual features contribute. In the following, we will first create an overview of published affective and stress classifiers, then we describe the methods used, followed by results, discussion and conclusion.

4.2 Affect and Stress Classification

Whereas there are multiple definitions of stress that differ in various subtleties, an often used definition is that of Lazarus & Folkman: "Psychological stress is a relationship between the person and the environment that is appraised by the person as taxing or exceeding his or her resources and endangering his or her well-being" (Lazarus and Folkman 1984). Stress can be measured through a variety of physiological signals, among which Skin Conductance (SC), Skin Temperature (ST), Electrocardiogram (ECG), Blood Volume Pulse (BVP), Blood Pressure (BP), Electroencephalogram (EEG) and Electromyogram (EMG) (Sharma and Gedeon 2012). Because emotions and other affective states can also be measured using these signals, it is worth positioning our work in the light of other affective classifications as well.

Table 4.1 shows shows a snapshot of ten affect classification studies from physiology. It can be seen that a variety of physiological modalities is used as input, various techniques are applied and a variety of target classes are used. Because the number of classes, number of participants, prior probability of classes and methods used for validation vary between these studies, their performance cannot be compared directly. Nevertheless one can observe that there is room for improvement in terms of performance, which ranges between 61% and 86%, with the majority of performances between 70 and 80%.

Table 4.2 shows a detailed overview of studies that specifically classify mental stress. We observe that the performances reported are slightly higher than those reported for other affective states (Table 4.1). We observe that most studies report 'ordinary' cross validation in which data of participants is shared over training and test set, only a limited number of studies report participant-wise cross validation results in which participants are strictly separated over training and test set (i.e., no data of test-participants is used for training). The latter is generally more difficult than the former, which becomes also apparent in the performances in Table 4.2, but does better reflect the generalization performance (i.e., performance of the method for unseen users).

Table 4.1: *Review of ten machine learning studies employing different physiological signals to recognize various affective states.*

Reference	Modalities[1]	Ss[2]	Feat.[3]	Technique	Targets	Perf[4]
Sinha and Parsons (1996)	M	27	18	LDA	2 emotions	86%
Picard et al. (2001)	C,E,R,M	1	40	LDA	8 emotions	81%
Kim et al. (2004)	C,E,S	175		SVM	3 emotions	73%
Lisetti and Nasoz (2004)	C,E,S	29		kNN, LDA, ANN	6 emotions	86%
Rani et al. (2006)	C,E,S,M,P	15	46	kNN, SVM, RT, BN	3 emotions	86%
Kim and André (2008)	C,E,M,R	3	110	LDA, EMDC[5]	4 emotions	79%
Chanel et al. (2009)	C,E,R	11	18		3 emotions	66%
	B		18720		3 emotions	73%
	B,C,E,R		18738		3 emotions	70%
Hosseini et al. (2010)	C,E,R	15	38	SVM	2 arousal levels	77%
	B	15	21	LDA, SVM	2 arousal levels	85%
van den Broek, Lisý, Janssen, Westerink, Schut and Tuinenbreijer (2010)	E,M	21	10	kNN, SVM, ANN	4 emotions	61%
Katsis et al. (2008)	C,E,M,R	10	15	SVM, ANFIS	4 affect states	79%

[1] Abbreviations used: B Brain activity (EEG); C Cardiovascular activity (e.g., ECG and BVP);E Electrodermal activity (EDA); M Electromyogram (EMG); P Blood pressure; R Respiration; S Skin temperature

[2] Number of subjects

[3] Number of features

[4] Performance (accuracy)

[5] A tailored ensemble of binary classifiers

Table 4.2: *Review of machine learning studies employing different physiological signals to recognize stress.*

Reference	Modalities[1]	Ss[2]	Technique	Targets	Val.[3]	Perf[4]
Healey and Picard (2005)	C,E,R,M	9	LDA	3-level	CV	97%
Zhai et al. (2005)	E,C,O	6	SVM (linear kernel)	2-class	CV	57%
			SVM (RBF kernel)			60%
			SVM (sigmoid kernel)			80%
Zhai and Barreto (2006)	E,C,O,S	32	SVM	2-class	CV	90%
	C,O,S					90%
	E,O,S					90%
	E,C,S					61%
	E,C,O					89%
Choi and Gutierrez-Osuna (2009)	C,R	3	unspecified	2-class	CV within pp	83%
					pp-wise CV	69%
Wijsman et al. (2011)	C,R,E,M	21	Linear Bayes	2-class	CV	78%
			Quadratic Bayes			78%
			kNN			76%
			Fisher's Least Square			79%
Giakoumis et al. (2013)	E	24	LDA	2-class	CV	83%
	C					74%
	E,C					95%
	E,C				pp-wise CV	86%

[1] Abbreviations used: B Brain activity (EEG); C Cardiovascular activity (e.g., ECG and BVP);E Electrodermal activity (EDA); M Electromyogram (EMG); O Ocular Response (e.g., Pupil diameter); P Blood pressure; R Respiration; S Skin temperature

[2] Number of subjects

[3] Type of validation

[4] Performance (accuracy)

The study of Healey and Picard (2005) provided an exceptionally high performance of 97%. It should, however, be noted that their study is limited in the number of participants used (13) as well as using only one task for each stress level. Therefore, the high performance they obtained is likely biased by the specific set of participants and might reflect distinctions between the tasks rather than the stress levels. In general, we observe that the number of participants used in the studies is relatively limited: the studies included data from 3 to 32 participants. In our study we gathered data from more participants to have a more representative set of participants. We repeated measurements in 15 sessions to introduce temporal effects and environmental changes in the dataset that happen in daily life and influence the physiological measurements. Furthermore, we use multiple stressful tasks to induce more variety to better represent stressful situations in daily life.

Sharma and Gedeon (2012) made an extensive inventory of various aspects of stress detection. They conclude that "Models developed to date that describe stress are quite simplistic. Generally, established techniques such as ANN and SVM have been used to model stress. Novel or more complex computational techniques are needed for stress models". We believe that the application of LVQ classifiers can be such a novel computational technique and help gain more direct insight into the stress classification challenge and thereby provide valuable input to develop models that describe stress.

4.3 Method

The experiment performed to obtain the data that will be used in the analysis that is subject of this work is further described in van Boxtel et al. (2012). The following sections describe the most important details. The current problem is defined as a binary classification problem in discerning stressful from relaxation episodes from human physiological signals. Stressful episodes were operationalized as various mentally demanding tasks, relaxation episodes were operationalised listening to favourite music. Human physiological signals entail the following modalities: ECG, GSR and RSP.

4.3.1 Participants

Participants were recruited by means of a website that explained the procedures involved in the research in great detail. A total number of 171 persons indicated on the website that they wanted to participate in the research. 110 persons either did not

follow up on our request, turned out to be unavailable at the time of the research, or decided to cancel their participation. The remaining 61 (20 male, 41 female) provided written informed consent. Their age ranged from 18 to 28 years (mean 21.2 years).

4.3.2 Design and procedure

Each participant returned 15 times within a period of 4 weeks for a session during which their physiology was measured. The sessions took place in a normal office room, in which each participant was seated in a comfortable reclining chair in front of a small table with a laptop on it. There were five such chairs and tables with laptops in the room, separated by wooden partitions, so that 5 participants could be trained at the same time by a single experimenter. The whole session was automated as much as possible. The experimenter supervised the sessions, and only took action in case something was wrong (usually bad electrode contacts, which were automatically signalled).

A training session on a particular day always consisted of the same sequence of tasks. After the signals were determined to be valid, a baseline measurement of five minutes rest with eyes opened was recorded, followed by 5 minutes with eyes closed. After that, 3 relaxation intervals of 8 minutes duration were interspersed by cognitive tasks lasting about 5 minutes each. The sequence of tasks are graphically represented in Figure 4.1. The (fixed) sequence was: Flanker task, relaxation 1, Stop-signal task, relaxation 2, Stroop task, relaxation 3, N-back task. During the relaxation intervals subconscious neuro-feedback was provided in three different ways, two of which are control conditions.

The interleaved task sequence was chosen for several reasons. First, it represents daily life stress, secondly it enhances changes in stress level which are particularly of interest for practical applications, and thirdly it provides a platform to test the relaxation effect of neuro-feedback.

Relaxation with Neuro-feedback

The participants were given a set of headphones that they used for listening to their favorite music. Participants could either bring their own music for that particular day on an MP3 player, or they could select that day's music from a playlist containing thousands of songs from various artists. There was no limitation to the kind of

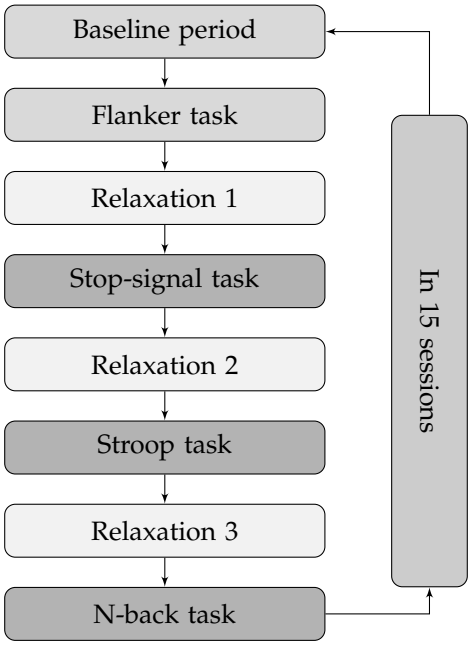

Figure 4.1: *Schematic outline of the experiment.*

music participants could listen to. Categories included genres like hard rock, easy listening and classical music.

As one part of this comprehensive study, the effects of neuro-feedback on relaxation were studied (van Boxtel et al. 2012). To that end, three conditions were used: alpha training and two types of controls, where one applies the same stimulation but at different (beta) frequencies that are not associated with relaxation and another control type where no stimulation is performed. Note that the stimulation was performed in a very subtle manner, as is described in the next paragraph, uses exactly the same setup over the three conditions, and has no direct effect on the peripheral physiological measurements taken (see Section 4.3.3).

The participants were randomly assigned to one of three groups: alpha training (A), random beta training (B), or control (C, music only); which was used for all sessions for this participant. Participants in group C listened to unaltered music, the music for the other two groups was altered by a high-pass filter of which the cut-off frequency was dynamically chosen. The cut-off frequency was adapted at real time based upon the frequency spectrum of the participants' EEG. To that end, the power in a target frequency range is calculated as relative to the total power

(i.e., the power in the range 4-35Hz). The higher this relative power, the lower the cut-off frequency was chosen. The resulting effect is that lower (relative) power in the target frequency bands causes the low frequencies of the music to be filtered out, while high (relative) power in the target frequency will pass the music without much change (in the lower music frequencies). The target frequency ranges in the EEG spectrum were chosen as follows: The range for group A was based upon the power of alpha waves (8-12Hz), and for group B it was based upon beta waves (a randomized 4Hz bin in the range 16-36Hz). The alpha training was expected to increase relaxation while the other two types were not expected to have any effect on relaxation. These expectations were confirmed by van Boxtel et al. (2012).

From the 61 participants, 50 completed all training sessions (and without technical problems). The participants were distributed over the three groups as follows: A (alpha training): $N = 18$ (12 female; mean age 20.7 ± 1.8 years); B (random beta training): $N = 12$ (9 female; mean age 20.6 ± 1.5 years); C (control, music only): $N = 20$ (15 female; mean age 21.0 ± 2.1 years). Further details on the neuro-feedback training can be found in van Boxtel et al. (2012), the present study focusses at the stress and relaxation aspects of this study.

The mentally demanding tasks are further detailed in the following, taken from the study protocol (Sitskoorn et al. 2009).

Stop-signal task

> " The stop-signal task basic choice reaction time task. A green triangle (0.050 of screen width) on a black background is presented on the computer screen. Subjects have to indicate as a fast as possible the direction of the triangle. For a triangle to the left, subjects press the most left button of a button box and when it points to the right, the most right button has to be pressed. In one third of the trials the green arrow becomes red for 100ms and no answer has to be given, as depicted in Figure 4.2. When subjects are able to stop their response, the next time the stop signal will be given 50ms later to make it more difficult. When subjects give a response despite the presence of a stop signal, the signal appears the next trial 50 ms earlier to make it easier for the subject to stop the response. The task starts with a stop signal delay time of 250 ms and depending on the reaction of the subject, the stop signal delay time changes. Logan and Cowan (1984) fitted performance on this task in a formal model. The present task will use staircase tracking of response rate to arrive at

50% of successfully stopped trials, which is an optimal value for estimating inhibitory efficiency (Stop Signal Reaction Time (SSRT)). After one trial was finished, a fixation cross of 0.004 of the screen width appeared between 1 and 2 seconds on the screen before the next trial started. ″ (Sitskoorn et al. 2009)

Figure 4.2: *Example of a Stop-signal task. The trial starts with a green arrow that depending on the subject's performance is green for a certain amount of time (at least 50 ms). After this time, the stop signal is initiated and the arrow becomes red for 100ms. This is followed by a green arrow that marks the end of the trial. It is the aim of the task that subjects do not give a response when the arrow becomes red.*

Stroop task

″ The computerized version of the Stroop Color Word Test (SCWT) (Zysset et al. 2001) is used as a measure of executive functioning. In the Stroop task, subjects have to indicate whether the meaning of a word is the same as the color of which another word is printed in. Both words are not presented at exactly the same time to make it more difficult for the subject. In our version of the Stroop task, both words are printed above each other and the first word is presented 150 ms before the other word. In the case of Figure 4.3, the word ″geel″ is presented 150 ms before the word ″rood″. Both words are visible during 500 ms. In this period, subjects have to indicate whether the color of the upper word is the same as the meaning of the lower word. This requires inhibition of the automatic response to read the color word (Hammes 1971). Hence, this test is considered a measure of 'disinhibition' and it generally has high reliability (Bouma et al. 1996).
In the example, the color of the word ″geel″ is red and the meaning of the lower word is red, so the trial is correct. When the color and color name correspond subjects press with their index finger the 'yes'-button, if not, they press the 'no'-button. Whether the right or left index finger

will be used for the yes and no response is counterbalanced between sessions. Congruent trials are trials in which the color of the upper word is the same as the meaning of this word. For example, the word ″Blue″ is written in blue ink and it means blue. When the color of the upper word is not the same as its meaning, the trial is called incongruent. In our experiment, four colors and the corresponding color names are used, namely red, yellow, blue and green. However, also the sign ″XXXX″ is used as an upper word. The expectation is that subjects will make fewer mistakes when ″XXXX″ is used as the upper word, because this word has no meaning and therefore subjects only have to deal with the color and not the meaning of the word. To keep between trials the attention of the subjects, a fixation cross with a variable duration between 1 and 2 seconds is presented on the computer screen. The duration of the fixation cross is variable to prevent a fixed rhythm of predicting and answering to the stimulus. ″ (Sitskoorn et al. 2009)

Figure 4.3: *An example of an incongruent matching trial in the Stroop task. The word ″geel″ means yellow but is written in red ink, so it is incongruent. The upper word is presented 150 ms before the lower word. Both words are visible during 500 ms. Between trials, a fixation cross with a duration between 1 and 2 seconds is presented on the computer screen.*

N-back task

″ The N-Back task is a working memory task, introduced by Kirchner (1958), and requires subjects to decide whether each stimulus in a sequence matches the one that appeared N items previously. For example in a 3-back task subjects have to decide whether a letter currently presented on the screen is the same as three letters earlier. Our version of the N-back task is a 2-back task, meaning that subjects should decide whether the letter on the screen was the same as two letters ago. The used test set consists of 8 letters, namely B, F, K, H, M, Q, R and X. We decided not to use vowels to prevent the formation of words, which are more easily remembered than single letters. Furthermore, we use letters who

are spatially different to be sure that when subjects make an error, it is caused by the difficulty of the task and not by confusion whether a letter was a V or W for example. Each letter will be presented for half a second on the computer screen. Between the end of a stimulus and the beginning of the next one, a fixation cross appeared on the screen. Except the 2-back trials, also lure trials were included in the task, such as depicted in Figure 4.4. These were trials in which the trial was 1-back or 3-back. When a trial was 2-back, subjects respond to the target by pressing the 'yes' button with their index finger. Whether the right or left index finger has to be used for the yes response is counterbalanced between sessions. " (Sitskoorn et al. 2009)

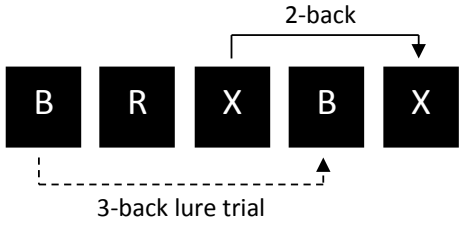

Figure 4.4: *An example of a sequence of letters in the N-back task. The last X matches the letter that was presented two items ago (X) and is therefore a 2-back. In this case subjects have to press the ijes-button on the button box, indicating that it was a 2-back. The letter B on the fourth position of the sequence is the same as the first one and is an example of a 3-back lure trial.*

Flanker task

" The Eriksen Flanker task (Eriksen and Eriksen 1974) is a basic choice reaction time task. In the task, five horizontally aligned arrows (size arrows: 0.050 of the screen width, space between arrows: 0.050 of the screen width) are presented on a 15 inch computer screen (resolution 1440 x 900 pixels, refresh rate 60 Hz) and subjects have to indicate the direction of the middle arrow (see Figure 4.5). Subjects can indicate this direction with a button box of which the most left button is pressed for an arrow pointing to the left and the most right button for an arrow pointing to the right. The two arrows on the left and right side of the

middle arrow are flanker arrows and presented 150 ms before the middle arrow. These arrows are meant to distract the subject. The four flanker arrows always point in the same direction to the left or right. In this way, two situations can occur, namely that the flankers point in the same direction as the middle arrow (congruent) or that the flankers point in the opposite direction of the middle arrow (incongruent). The middle arrow with flankers will be present for 500 ms. After this period, a fixation cross (0.004 of screen width) appears at the same position as the middle arrow, namely in the center of the screen. To prevent that subjects learn when the next trial will start, the duration of the fixation cross will vary between 1 and 2 seconds. " (Sitskoorn et al. 2009)

Figure 4.5: *Stimuli used in the Flanker task. The two arrows on the left and right side of the middle arrow are the flanker arrows and presented 150 ms before the middle arrow appears. The arrows are white and presented on a black background. Subjects have to indicate the direction of the middle arrow. a) Congruent situation. The flanker arrows point in the same direction as the middle arrow. b) Incongruent situation. The flanker arrows point into the opposite direction of the middle arrow.*

For the classification analysis described in the Section 4.3.4, we selected the data gathered during the three relaxation tasks and the Stop-signal, Stroop and N-back tasks (as mentally stressful tasks). We did not include the Flanker task in the analysis as it turned out that participants were able to master the Flanker task very well after only a few attempts, thereby strongly reducing the mental stressfulness of the task in subsequent sessions. After each task the participants were asked to rate their level of stress vs relaxation on a visual analogue scale. Effectiveness of the induction of stress vs relaxation was tested by applying an ANOVA with repeated measures to these reported levels of stress.

4.3.3 Measurements

GSR was recorded from the left index finger, ECG was recorded from an electrode placed on the left wrist, and RSP was measured using a chest belt with stretch sensor. The signals were sampled at a rate of 1024 Hz (ECG), and 256 Hz (GSR, and RSP) by a 24 bit A/D converter on a Nexus-10 portable device (MindMedia B.V.,

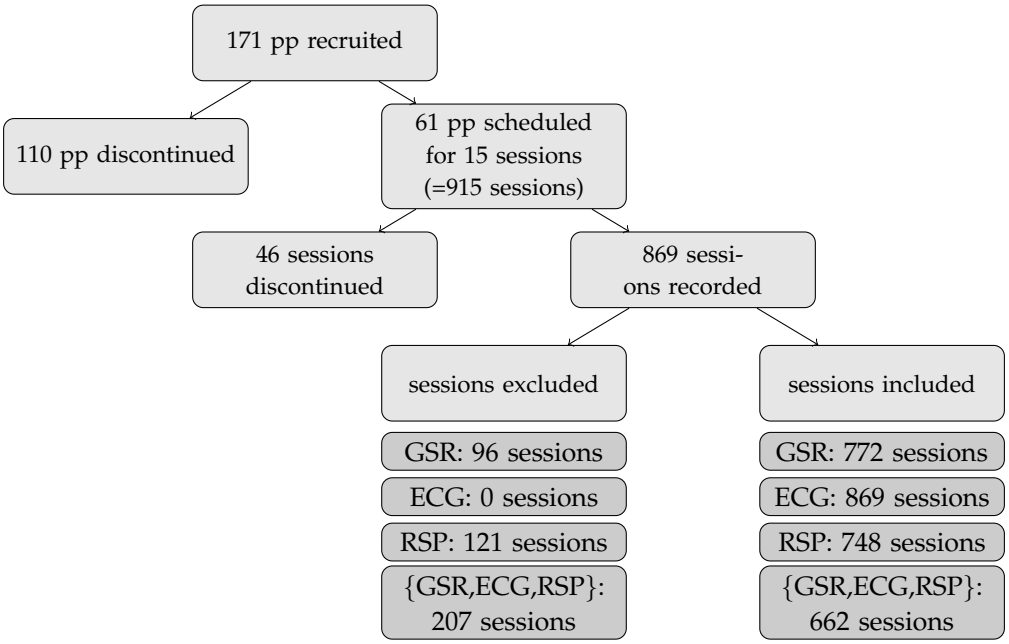

Figure 4.6: *Schematic outline of the data selection/exclusion process. Left branches show exclusion, and right branches inclusion of sessions.*

The Netherlands).

For each participant there were 15 sessions scheduled totalling 915 ($= 61 * 15$) sessions, of which 46 were discontinued due to technical problems or unconformity of participants, yielding 869 sessions. Due to bad signal quality (e.g., signals out of range of the measuring equipment) we further excluded, 96, and 121 sessions for GSR and RSP respectively, resulting in 772, and 748 sessions from analyses for these signals. No ECG sessions needed to be excluded. In total 662 sessions contained valid signals for all modalities. Figure 4.6 depicts the data selection (or exclusion) schematically.

Preprocessing & Feature extraction

The steps taken during preprocessing and feature extraction are schematically depicted in Figure 4.7. As a first step of preprocessing, the signals were downsampled to 512 Hz (ECG), and 128 Hz (GSR, and RSP). Subsequently, signals were analyzed

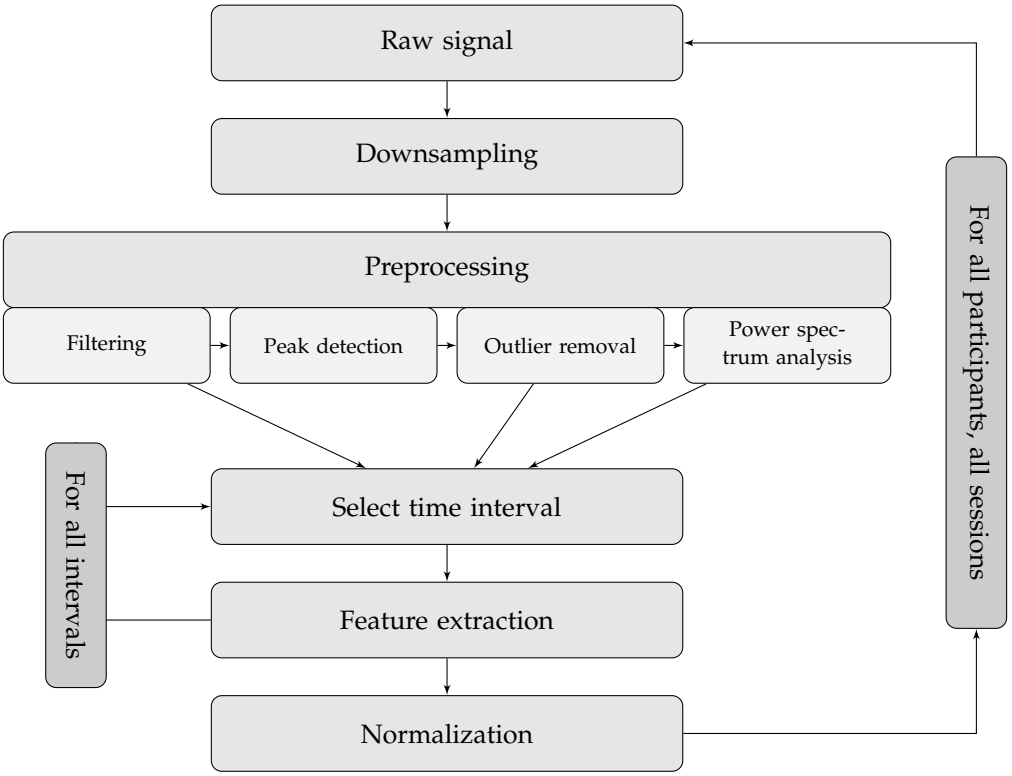

Figure 4.7: *Schematic overview of the feature extraction process.*

through the following dedicated preprocessing methods:

ECG preprocessing consisted of the following steps (as outlined in de Waele et al. (2009)): R-peak detection, IBI outlier removal, and Heart Rate Variability (HRV) analysis. R-peak detection was performed using a pattern matching technique (Poor 1994). The resulting intervals between the R-peaks, called Inter-Beat Interval (IBI), are filtered for outliers by using a sliding window histogram. In order to estimate frequency domain HRV features, an Autoregressive-Moving Average (ARMA) time series model was used to derive power in the frequency bands defined in the HRV guidelines paper (Malik et al. 1996), ranging from 0.04 to 0.15 Hz, which is known to vary with parasympathetic nervous system activity (Grossman and Taylor 2007). Next to the frequency domain HRV features, a variety of time domain HRV features is calculated, given that no 'golden standard' for HRV has been defined (Allen et al. 2007), as well as several features based on plain IBIs.

GSR was preprocessed using the SCRGauge method described in Kohlish (1992) which first subsamples the GSR signal to 1 Hz, uses cubic splines interpolation followed by a dedicated local maximum detection which is triggered by exceeding a certain gradient. Backward and forward searches are subsequently applied to detect the onset of Skin Conductances Responses (SCRs), and half recovery times. The raw GSR signal was used to derive several Skin Conductance Level (SCL) features from, the extracted SCRs to derive SCR features from, and from the residual signal that resides after subtracting SCRs from the raw signal, using the technique described in de Vries and van der Zwaag (2010) we derived features that represent purely the tonic part of GSR.

RSP signals were first lowpass filtered (cut-off 0.5Hz) and then analyzed for individual breaths. Using a localized min/max filter (Lemire 2006), local minima and maxima are detected. When found in the right order, they characterize a single breath. Based upon the distribution of identified breath amplitudes in a signal, too small or too large breaths (outliers) are removed. After this preprocessing the RSP signal is characterized by a sequence of breaths similar to the IBI signal for ECG.

All features have been calculated over equal length time intervals in order to avoid bias in duration dependent features (such as standard deviations) towards certain tasks. To this end, the first 5 minutes (which is the minimal duration of tasks) of measured signals from each task was taken to derived the feature values. A complete overview of extracted features can be found in Table 4.3. The specific features have been chosen such that they express the dynamics known to be relevant (Dawson et al. 2000, Stern et al. 2001, Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology 1996) as they are modulated by the Autonomous Nervous System (ANS) that responds to stress. From this large set of features, we compiled a subset of features representing the most often used features in literature (inspired by the list in van den Broek et al. (2009)). They are marked with an asterisk in Table 4.3. In order to combine the data gathered from the different physiological signals to be used by a single classifier, we applied feature level fusion.

As highlighted in van den Broek, van der Zwaag, Healey, Janssen and Westerink (2010), there are many different techniques for normalization. In addition to the choice of which correction formula to use, the choice in defining the baseline period, there is also the choice of correcting on signal or feature level. The aim of normalization is to reduce the variance that occurs due to differences in physiology

between participants, but also the long-term changes in physiology over time within participants, e.g., due to differences in physical fitness or the environment (such as temperature and humidity) (Boucsein 1992, Boucsein 2012). We have chosen for z-correction, a technique that compensates both for baseline level and variation, and is not too sensitive to outliers. Rather than applying the correction to the raw signals (which would only make sense for the skin conductance level), we apply it to the features derived, and we use the entire recording (all tasks) as reference signal, as suggested for e.g., SCR amplitude by Boucsein (Boucsein 1992, Boucsein 2012). Hence, after computing the features per task, we applied z-correction ($x_{corr} = \frac{x-\mu}{\sigma}$) to compensate for differences in physiological baselines between people and sessions. In this formula, μ represents the mean of a feature's values over all tasks within a single session (for a single participant), and σ the respective standard deviation.

Table 4.3: *Features extracted from the raw and preprocessed signals.*

ECG	IBI min	minimal IBI
	IBI max	maximal IBI
	IBI mean *	mean IBI
	IBI std *	standard deviation of IBIs, also referred to as SDNN
	IBI amp	amplitude of IBIs (max-min)
	IBI power VLF *	power of IBIs in very low frequency band ($0 - 0.04$ Hz)
	IBI power LF *	power of IBIs in low frequency band ($0.04 - 0.15$ Hz)
	IBI power HF *	power of IBIs in high frequency band ($0.15 - 0.4$ Hz)
	IBI power VHF	power of IBIs in very high frequency band ($0.4 - 1$ Hz)
	IBI power LH *	ratio between IBI power LF and HF
	IBI RMSSD *	root mean square of successive differences of IBIs
	IBI PNN50	proportion of IBIs > 50 ms
	IBI SDSD	standard deviation of successive differences of IBIs
GSR	SCL mean *	mean SCL
	SCL std *	standard deviation of SCL
	SCL grad	gradient of SCL (estimated by best linear fit)
	SCL min	minimal SCL
	SCL max	maximal SCL
	SCR freq *	number of SCRs per second
	SCR max amp	maximal amplitude of SCRs
	SCR mean amp *	mean amplitude of SCRs
	SCR sum amp	sum of amplitudes of SCRs
	SCR mean rise time *	mean rise time of SCRs
	SCR mean rec time *	mean half recovery time of SCRs
	SCR mean rise rec	mean ratio of rise time and half recovery time of SCRs
	SCR mean rise amp	mean ratio of rise time and amplitude of SCRs
	SCR mean rec amp	mean ratio of half recovery time and amplitude of SCRs
	SCRc SCL mean	mean SCL after correcting for SCRs
	SCRc SCL std	standard deviation of SCL after correcting for SCRs
	SCRc SCL grad	gradient of SCL after correcting for SCRs (estimated by best linear fit
	SCRc SCL min	minimal SCL after correcting for SCRs
	SCRc SCL max	maximal SCL after correcting for SCRs
RSP	mean rate *	mean respiration rate
	median rate	median respiration rate
	mean amp *	mean amplitude of respirations
	mean inh time	mean inhalation time
	mean exh time	mean exhalation time
	mean cycle	mean respiration time
	mean duty cycle	ratio between mean inhalation time and cycle
	mean inh exh	ratio between mean inhalation and exhalation time

* This feature is included in the commonly used set of features representing all modalities.

4.3.4 Classification analysis

In order to answer our research question of discerning stress from relaxation using physiology as input, we applied a selection of classifiers and further optimized their parameter settings using data from the individual physiological modalities (GSR, ECG and RSP) as well as the combined multi-modal dataset. Finally, we used the trained LVQ classifiers to derive which features were most influential in distinguishing stress from relaxation.

Learning Vector Quantization (LVQ) comprises a family of classifiers that is of open box nature, that is, they provide direct insight into the information learned by the classifier. LVQ, initially proposed by Kohonen (1990), defines prototypes $\mathbf{w}_T$ in the same (mathematical) space as the data (samples ξ) to represent the classes. These prototypes are directly interpretable as they show characteristics of classes in terms of the features chosen. During training, samples are presented sequentially, and for each sample the closest prototype(s) are updated by moving them towards or away from the presented sample. Several variants have been proposed, amongst which Robust Soft Learning Vector Quantization (RSLVQ) (Seo and Obermayer 2003), which introduces soft prototype assignments which act similarly to a soft window around the decision boundary (Witoelar et al. 2011), and Generalized Matrix Learning Vector Quantization (GMLVQ) (Schneider et al. 2009a), which introduces a relevance matrix Λ that is trained along with the prototypes. $\Lambda = \Omega^\top \Omega$, with Ω of size $M \times N$, is used to adapt the distance measure used by LVQ according to Equation 2.17. We have trained GMLVQ both with $2 \times N$ and $5 \times N$ sized matrices Ω, but since we observed identical performances, we will only report results of $2 \times N$ sized Ω. We will present results for RSLVQ and GMLVQ using one prototype per class as using more prototypes per class did not improve the results. In addition, we apply SVM (Vapnik 1998), a very popular technique in this domain of biomedical engineering. Next to linear SVM, which will be reported in the results, we also applied SVM with an RBF kernel, which however, did not improve upon the results.

Cross validation

In order to estimate generalization performance, we employed a cross validation scheme. Because physiological data shows large variation between participants (Gale and Edwards 1983, Boucsein 1992, Boucsein 2012), the most applicable, but also most challenging classification task is to separate training and test data not only per sample, but per subject. Hence we used 10 fold participant-wise cross validation, which divides the set of participants in tenths and repeatedly uses data from

90% of participants for training and the rest for testing. The results reported are means and standard deviations over 10×10-fold participant-wise cross validations. In addition to the participant-wise cross validation, we also performed 'ordinary' cross validation in which data of single participants can be split over both training and test set, thereby leaking some information from 'training participants' to the test set.

4.4 Results

Per participant, we have 15 sessions comprising 3 repetitions of 2 tasks, one operating in a stress condition and one operating in a relaxation condition. As dependent variable, we asked participants to report stress level using a visual analogue scale from zero to one (mean for relax condition: 0.29; for stressful condition: 0.38). It is evident that mean reported stress levels are derived from same participants measured in different sessions, repetitions and tasks, not from different participants. This refers to a within-subject or repeated measure design for statistical analysis. To demonstrate the induction of stress by means of mental and relaxation tasks, the aim of the analysis is to reject the null hypothesis of no difference in mean 'reported stress level' between stress conditioned tasks and relaxation conditioned tasks. We conducted an ANOVA with repeated measures with 'reported stress level' as dependent variable, and session (15), repetitions (3) and tasks (2) as within-subject independent variables. Missing values in reported stress levels were dealt with by means of case-based exclusion. We found a significant main effect for tasks ($F(1, 36) = 38.3$, $p < 0.001$) allowing us to reject the null hypothesis of no difference in reported stress levels.

Tables 4.4 and 4.5 show the means and standard deviations per feature per task. They indicate, e.g., that the number of SCRs observed in the relaxation tasks is generally lower than in the mentally stressful tasks, as is the average SCL, respiration rate and amplitude. The heart rate variability, measured e.g. through IBI RMSSD, is generally higher in the stressful tasks. Within the mentally stressful or relaxation tasks the three subtasks show similar feature values.

The classification results of the participant-wise cross validation are shown in Table 4.6 when using data from single modalities, and Table 4.7 when combining data from multiple modalities. It can be observed that the performances are very similar over different classification techniques. Comparing the single modalities, the classifiers perform best on ECG, followed by GSR and RSP. Combining features from

Table 4.4: *Statistics (means and standard deviations) per feature per task of non-normalized data. See Table 4.3 for explanation of the feature names used.*

Feature name | Task	Relax 1	Relax 2	Relax 3	N-back	Stop-signal	Stroop
SCR freq	0.029 ± 0.034	0.029 ± 0.036	0.031 ± 0.036	0.057 ± 0.057	0.059 ± 0.059	0.062 ± 0.063
SCL mean	2.043 ± 1.398	2.122 ± 1.454	2.184 ± 1.460	2.447 ± 1.537	2.304 ± 1.518	2.416 ± 1.554
SCL std	0.248 ± 0.201	0.244 ± 0.197	0.242 ± 0.193	0.215 ± 0.184	0.240 ± 0.209	0.237 ± 0.195
RSP rate	15.432 ± 3.827	15.080 ± 3.867	14.951 ± 3.877	16.303 ± 3.700	16.488 ± 3.391	16.399 ± 3.580
RSP mean amp	6.367 ± 4.755	6.452 ± 5.016	6.419 ± 4.978	6.947 ± 5.124	6.781 ± 5.114	7.048 ± 5.766
IBI mean	0.825 ± 0.112	0.839 ± 0.113	0.853 ± 0.116	0.842 ± 0.115	0.835 ± 0.113	0.836 ± 0.114
IBI RMSSD	0.053 ± 0.034	0.057 ± 0.039	0.060 ± 0.038	0.073 ± 0.056	0.071 ± 0.052	0.071 ± 0.050

Table 4.5: *Statistics (means and standard deviations) per feature per task of normalized data. See Table 4.3 for explanation of the feature names used.*

Feature name | Task	Relax 1	Relax 2	Relax 3	N-back	Stop-signal	Stroop
SCR freq	-0.343 ± 0.712	-0.269 ± 0.845	-0.199 ± 0.843	0.433 ± 0.977	0.380 ± 0.848	0.441 ± 0.882
SCL mean	-0.202 ± 0.645	-0.048 ± 0.727	0.096 ± 0.803	0.647 ± 0.889	0.329 ± 0.650	0.563 ± 0.705
SCL std	0.069 ± 0.887	0.035 ± 0.891	0.031 ± 0.920	-0.205 ± 0.876	0.009 ± 0.854	0.021 ± 0.830
RSP rate	-0.279 ± 0.813	-0.449 ± 0.825	-0.522 ± 0.912	0.065 ± 1.050	0.197 ± 0.823	0.164 ± 0.860
RSP mean amp	-0.149 ± 0.874	-0.164 ± 0.851	-0.123 ± 1.007	0.139 ± 1.035	0.059 ± 0.849	0.131 ± 0.923
IBI mean	-0.065 ± 0.745	0.237 ± 0.793	0.627 ± 0.874	0.348 ± 1.016	0.251 ± 0.716	0.201 ± 0.809
IBI RMSSD	-0.238 ± 0.825	-0.114 ± 0.814	0.187 ± 0.917	0.585 ± 0.912	0.463 ± 0.751	0.455 ± 0.805

the three modalities improves performance, and adds an additional 3.5 percentage points on top of the performance obtained using only ECG features. This, however, only when including more than just the set of commonly-used features. For the richest data set covering all features, GMLVQ performs best with just under 87% accuracy. The corresponding Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) is 0.95. Using 'ordinary' cross validation, performance is slightly higher with accuracies up to 88%.

The training performance is presented in Tables 4.8 and 4.9, again for uni- and multimodal input, respectively. It can be observed that the training performances for LVQ and linear SVM are only slightly higher than their respective test performances (Tables 4.6 and 4.7), while for the ANN and RBF SVM there is a larger difference in performance. We also verified whether parameter settings could be optimized, w.r.t. generalization performance, using only training data (and training performance) and found that this is possible on the present dataset, within two percentage points of the overall optimum, for all classifiers except kNN and ANN.

Table 4.6: *Test performance of 10x10 fold participant-wise cross validation on normalized data using single-modalities. Accuracy and AUC are listed as mean and standard deviation over the 10 repetitions of the 10-fold cross validation.*

	RSP			GSR			ECG		
Method	Hyper-parameter	Accuracy	AUC	Hyper-parameter	Accuracy	AUC	Hyper-parameter	Accuracy	AUC
Baseline		50.0%	0.5		51.4%	0.5		50.0%	0.5
kNN	$k = 11$	$69.0\% \pm 0.3\%$	$NaN \pm NaN$	$k = 11$	$71.9\% \pm 0.2\%$	$NaN \pm NaN$	$k = 11$	$81.7\% \pm 0.2\%$	$NaN \pm NaN$
ANN	$N_{hidden} = 3$	$70.4\% \pm 0.5\%$	0.748 ± 0.091	$N_{hidden} = 3$	$73.6\% \pm 0.6\%$	0.813 ± 0.070	$N_{hidden} = 5$	$84.1\% \pm 0.2\%$	0.913 ± 0.094
SVM - linear	$C = 1$	$71.0\% \pm 0.2\%$	0.748 ± 0.089	$C = 0.1$	$74.8\% \pm 0.3\%$	0.814 ± 0.062	$C = 0.1$	$83.2\% \pm 0.1\%$	0.897 ± 0.116
SVM - RBF	$C = 1$	$69.7\% \pm 0.5\%$	0.745 ± 0.068	$C = 10$	$69.7\% \pm 0.3\%$	0.775 ± 0.075	$C = 1$	$81.6\% \pm 0.3\%$	0.882 ± 0.106
Means		$65.7\% \pm 0.1\%$	0.684 ± 0.117		$71.3\% \pm 0.1\%$	0.778 ± 0.064		$76.8\% \pm 0.2\%$	0.846 ± 0.130
GLVQ		$67.7\% \pm 0.2\%$	0.709 ± 0.115		$71.6\% \pm 0.3\%$	0.783 ± 0.064		$77.2\% \pm 0.2\%$	0.848 ± 0.136
RSLVQ	$v_{soft} = 0.5$	$70.9\% \pm 0.3\%$	0.746 ± 0.091	$v_{soft} = 1$	$74.8\% \pm 0.2\%$	0.812 ± 0.062	$v_{soft} = 1$	$83.1\% \pm 0.1\%$	0.895 ± 0.109
GRLVQ		$68.0\% \pm 0.4\%$	0.707 ± 0.121		$66.6\% \pm 1.9\%$	0.715 ± 0.107		$70.2\% \pm 1.8\%$	0.721 ± 0.151
GMLVQ		$71.0\% \pm 0.3\%$	0.744 ± 0.091		$74.4\% \pm 0.3\%$	0.810 ± 0.058		$83.4\% \pm 0.2\%$	0.895 ± 0.115

Table 4.7: *Test performance of 10x10 fold participant-wise cross validation on normalized data combining multiple modalities. Accuracy and AUC are listed as mean and standard deviation over the 10 repetitions of the 10-fold cross validation.*

	Multi-selection			Multi-all		
Method	Hyper-parameter	Accuracy	AUC	Hyper-parameter	Accuracy	AUC
Baseline		51.3%	0.5		51.3%	0.5
kNN	$k = 11$	$79.1\% \pm 0.2\%$	-	$k = 11$	$81.7\% \pm 0.3\%$	-
ANN	$N_{hidden} = 3$	$81.6\% \pm 0.4\%$	0.848 ± 0.126	$N_{hidden} = 5$	$85.4\% \pm 0.5\%$	0.936 ± 0.017
SVM - linear	$C = 0.1$	$81.4\% \pm 0.3\%$	0.848 ± 0.138	$C = 0.1$	$86.6\% \pm 0.2\%$	0.954 ± 0.015
SVM - RBF	$C = 10$	$73.8\% \pm 0.3\%$	0.792 ± 0.138	$C = 0.001$	$51.3\% \pm 0.0\%$	0.849 ± 0.034
Means		$78.9\% \pm 0.2\%$	0.821 ± 0.132		$80.2\% \pm 0.2\%$	0.915 ± 0.029
GLVQ		$78.9\% \pm 0.2\%$	0.818 ± 0.129		$80.6\% \pm 0.2\%$	0.913 ± 0.029
RSLVQ	$v_{soft} = 1$	$81.6\% \pm 0.3\%$	0.847 ± 0.133	$v_{soft} = 1$	$86.6\% \pm 0.2\%$	0.950 ± 0.016
GRLVQ		$64.7\% \pm 2.0\%$	0.641 ± 0.130		$72.2\% \pm 1.1\%$	0.773 ± 0.080
GMLVQ		$81.6\% \pm 0.2\%$	0.845 ± 0.135		$86.7\% \pm 0.2\%$	0.951 ± 0.016

Table 4.8: *Training performance of 10x10 fold participant-wise cross validation on normalized data using single-modalities. Accuracy and AUC are listed as mean and standard deviation over the 10 repetitions of the 10-fold cross validation.*

	RSP			GSR			ECG		
Method	Hyper-parameter	Accuracy	AUC	Hyper-parameter	Accuracy	AUC	Hyper-parameter	Accuracy	AUC
Baseline		51.4%	0.5		51.4%	0.5		50.0%	0.5
kNN	$k = 11$	$75.3\% \pm 0.1\%$	-	$k = 11$	$77.1\% \pm 0.1\%$	-	$k = 11$	$84.8\% \pm 0.0\%$	-
ANN	$N_{hidden} = 3$	$72.7\% \pm 0.1\%$	0.798 ± 0.009	$N_{hidden} = 3$	$77.3\% \pm 0.2\%$	0.846 ± 0.007	$N_{hidden} = 5$	$85.7\% \pm 0.1\%$	0.929 ± 0.003
SVM - linear	$C = 1$	$71.9\% \pm 0.0\%$	0.783 ± 0.009	$C = 0.1$	$75.6\% \pm 0.0\%$	0.818 ± 0.006	$C = 0.1$	$83.7\% \pm 0.0\%$	0.903 ± 0.004
SVM - RBF	$C = 1$	$78.4\% \pm 0.1\%$	0.869 ± 0.005	$C = 10$	$100.0\% \pm 0.0\%$	1.000 ± 0.000	$C = 1$	$97.2\% \pm 0.0\%$	0.996 ± 0.000
Means		$66.1\% \pm 0.0\%$	0.712 ± 0.011		$71.7\% \pm 0.0\%$	0.777 ± 0.007		$77.3\% \pm 0.0\%$	0.851 ± 0.005
GLVQ		$68.3\% \pm 0.1\%$	0.738 ± 0.009		$71.9\% \pm 0.1\%$	0.780 ± 0.009		$77.5\% \pm 0.1\%$	0.850 ± 0.006
RSLVQ	$v_{soft} = 0.5$	$71.7\% \pm 0.1\%$	0.782 ± 0.009	$v_{soft} = 1$	$75.6\% \pm 0.0\%$	0.816 ± 0.006	$v_{soft} = 1$	$83.6\% \pm 0.0\%$	0.899 ± 0.004
GRLVQ		$67.9\% \pm 0.2\%$	0.717 ± 0.013		$67.1\% \pm 0.8\%$	0.689 ± 0.062		$70.3\% \pm 1.3\%$	0.720 ± 0.081
GMLVQ		$71.9\% \pm 0.1\%$	0.778 ± 0.011		$76.0\% \pm 0.1\%$	0.816 ± 0.006		$83.8\% \pm 0.0\%$	0.901 ± 0.004

Table 4.9: *Training performance of 10x10 fold participant-wise cross validation on normalized data combining multiple modalities. Accuracy and AUC are listed as mean and standard deviation over the 10 repetitions of the 10-fold cross validation.*

	Multi-selection			Multi-all		
Method	Hyper-parameter	Accuracy	AUC	Hyper-parameter	Accuracy	AUC
Baseline		51.3%	0.5		51.3%	0.5
kNN	$k = 11$	$83.4\% \pm 0.0\%$	-	$k = 11$	$85.8\% \pm 0.1\%$	-
ANN	$N_{hidden} = 3$	$84.0\% \pm 0.1\%$	0.918 ± 0.005	$N_{hidden} = 5$	$92.0\% \pm 0.3\%$	0.971 ± 0.003
SVM - linear	$C = 0.1$	$82.9\% \pm 0.0\%$	0.906 ± 0.006	$C = 0.1$	$88.2\% \pm 0.0\%$	0.947 ± 0.002
SVM - RBF	$C = 10$	$100.0\% \pm 0.0\%$	1.000 ± 0.000	$C = 0.001$	$51.3\% \pm 0.0\%$	1.000 ± 0.000
Means		$79.6\% \pm 0.0\%$	0.878 ± 0.006		$81.0\% \pm 0.0\%$	0.886 ± 0.003
GLVQ		$79.6\% \pm 0.0\%$	0.876 ± 0.007		$81.5\% \pm 0.0\%$	0.885 ± 0.003
RSLVQ	$v_{soft} = 1$	$82.7\% \pm 0.0\%$	0.904 ± 0.005	$v_{soft} = 1$	$88.2\% \pm 0.0\%$	0.943 ± 0.002
GRLVQ		$66.3\% \pm 1.1\%$	0.710 ± 0.075		$72.9\% \pm 0.5\%$	0.774 ± 0.021
GMLVQ		$83.0\% \pm 0.0\%$	0.903 ± 0.005		$88.4\% \pm 0.1\%$	0.943 ± 0.002

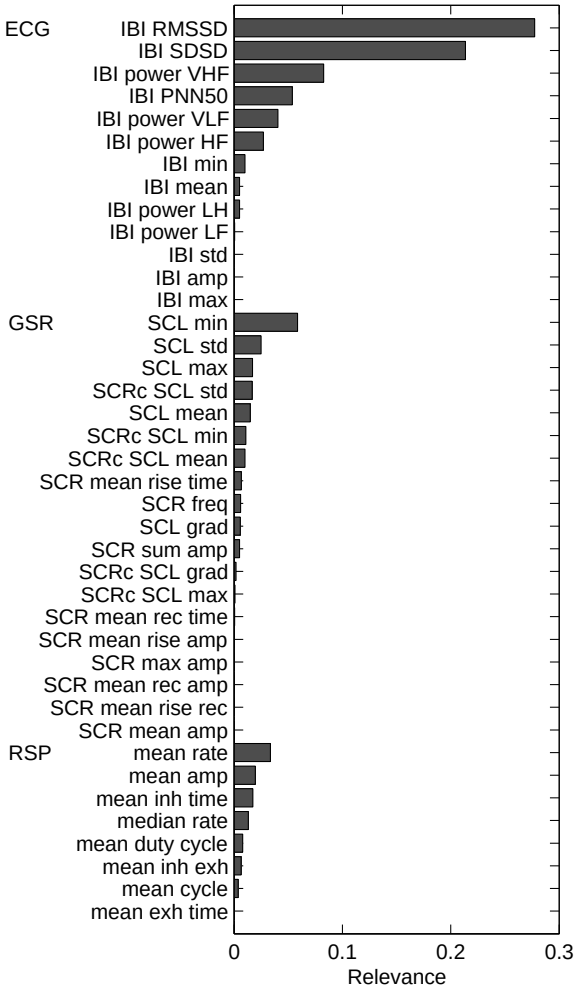

Figure 4.8: *Relevances trained by GMLVQ. For explanation of the feature names used, see Table 4.3.*

The diagonal elements of the relevance matrix trained by GMLVQ indicate relevance of individual features to the classifier's decisions. Figure 4.8 shows these relevances and indicates that most informative features come from the ECG modality, followed by GSR and RSP. The most important individual contributions come from the time domain HRV measures SDSD and RMSSD, followed by the frequency domain HRV measured by the power in the VHF range. The most influential GSR feature is the minimum SCL and for RSP the mean rate.

4.5 Discussion

The classification performances for uni-modal stress classification range from 71% for RSP, to 83.4% for ECG indicating that most information can be found in the ECG signal. By combining modalities the performance is further increased to 86.7% in participant-wise cross-validation. Using 'ordinary' cross-validation we obtained an accuracy of 87.7%. Both LVQ methods we employed (RSLVQ and GMLVQ) performed well and slightly outperformed the popular SVM, that we included for reference.

From the training performances we observe that ANN and RBF SVM have the tendency to overtrain and thereby have a poor generalization performance. The parameter settings of LVQ variants as well as linear SVM could be optimized using training performances, thereby enabling to build a classifier solely on training data without the need for a test or validation set while yielding good generalization performance.

With these accuracies, our methods outperform the affective classifiers that are listed in Table 4.1. They also outperform the participant-wise validated stress-classifiers in Table 4.2 as well as most others using other cross validation schemes. In comparison to other studies, we used data obtained from a larger number of participants, and also repeated measurements for every participant in 15 sessions spread over several weeks. Thereby, we used a more representative sample of participants and obtained reliable estimates of generalization performance.

The most important feature was found to be the RMSSD. Although it reflects high frequency modulation of heart rate that, in general, is affected by Respiratory Sinus Arrhythmia (RSA), RMSSD has been shown to be unaffected of breathing (Penttilä et al. 2001). The second most influential measure was the related SDSD. The PNN50 that correlates with RMSSD (Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology 1996), was also identified as quite relevant. It is worth noting that the two most influential features (RMSSD and SDSD) have been found earlier to be most reliable measures for short term intervals (McNames and Aboy 2006), i.e., in the order of 5 minutes, which reflects the measurement time in our experiments. The importance of various HRV measures for distinguishing can be explained by the fact that they reflect parasym-

pathetic (HF power, RMSSD) and sympathetic (LF power, SDNN) nervous system responses which are known to relate to the fight-or-flight response and dampening response, respectively (Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology 1996).

Out of the frequency domain HRV measures VHF power we found most influential. Many studies that include cardiac activity as measure for stress only use features from the lower frequency (up to HF) ranges and do not usually consider frequencies above 0.4Hz. This might be due to various reasons. First, the mechanisms that affect VHF power are not well understood. Second, higher frequency ranges cannot be measured reliably through the most commonly used modality BVP which has less sharp peaks, thereby not allowing for a very accurate detection of heart beats which is reflect particularly in inaccuracy in the higher frequencies of HRV. Our use of ECG enabled the reliable use of VHF power as a measure of stress.

We also inspected the prototypes trained by GMLVQ to verify that the stress prototype, as compared to the relax prototype, is characterized by higher heart rate and generally higher HRV values with the exception of RMSSD. Especially for HRV there are varying results published (Mathewson et al. 2010, Wright et al. 2007, Vuksanović and Gal 2007), which is confirmed by Berntson and Caciopo, who observed that "it is clear that no single pattern of autonomic adjustments and associated changes in heart rate variability will apply universally across different stressors" (Berntson and Cacioppo 2004). In our study we included three different stressors to induce stressful situations, thereby creating more robustness against this effect. Further, the stress prototype is characterized by more SCRs, higher maximum SCL and faster, though deeper, breathing. These findings are in line with observations made by others (Boucsein 2012, Houtveen et al. 2002).

4.6 Conclusion

We have successfully built classifiers of stress from three physiological modalities and observed that the cardiac activity made the strongest uni-modal classifier with over 83% accuracy. Combining the three modalities into a multi-modal classifier improved performance further up to 88% accuracy. By using data from a large sample of participants and repeated sessions we ensured good generalizability to unseen users. The LVQ techniques slightly outperformed well-known techniques such as SVM. These open-box methods allowed us to observe the most important features

for stress detection. These were the time domain HRV measures RMSSD and SDSD. The third most important features was found to be very high-frequency HRV from ECG. Most other studies use BVP to measure cardiac activity, however that does not allow for accurate VHF HRV measurements. Therefore it might be advisable for methods that aim at stress detection to use ECG rather than BVP as measurement modality of cardiac activity.

The classifiers built and the knowledge gained on important features for the distinction between stress and relaxation using physiological parameters have brought us one step closer to the realization of a system that can monitor physiology during the day and help its users to monitor their stressful moments during the day. In case a certain quota has been reached or a stressful period reaches a certain duration such a system could trigger the user and offer a means to relief the stress, e.g., a paced breathing exercise (Westerink et al. 2014).

While we have setup our experiments such that they represent daily life as well as possible, the measurements were taken during semi-lab circumstances. Future work should look into the application of the developed classifiers in daily-life measurements and observe their performance. This brings the challenge of reliable ground truth measurements, however this might become more and more feasible with the rapid development of various technologies such as Google Glass (Google 2014) that can capture context. It would be interesting to expand the classifier to also be able to classify other affective phenomena such as emotions and moods.

Material based on: de Vries, J. J. G., Pauws, S. C. and Biehl, M.: submitted, Facial expression recognition using learning vector quantization, Pattern Recognition

Hoofdstuk 5
EMOTION FROM A FACIAL PERSPECTIVE

Abstract

The detection of emotions from facial video or images has been topic of research for several years, nevertheless the set of applied classification techniques seems limited to a few popular methods. Benchmark datasets facilitate direct comparison of methods. We used one such dataset, the Cohn-Kanade database, to build classifiers for facial expression recognition based upon Local Binary Patterns (LBP) features. We are interested in the application of Learning Vector Quantization (LVQ) classifiers to this classification task. These prototype-based classifiers allow to inspect of prototypical features of the emotion classes, are conceptually intuitive and quick to train. For comparison we also consider Support Vector Machine (SVM) and observe that LVQ performances exceed those reported in literature for methods based upon LBP features and are amongst the overall top performing methods. Most prominent features were found to originate, primarily, from the mouth region and eye regions. Finally, we explored the specific LBP features that are found most influential within these regions.

5.1 Introduction

In human history, facial expressions have grown as important element of inter-human communication. Especially since mankind developed social emotions, estimated to date back to 2 million years ago at the time of the early *Homo Erectus* (Dubreuil 2010), the face became the primary means for emotional expression. Many applications, especially in human-computer interaction can benefit from facial expression recognition of its users (Peter and Beale 2008), ranging from affective content selection to adaptive system behavior to the affective state of the user. For many of such applications, users have restricted range of motion, enabling the measurement of affect through unobtrusive measurement using video or photo cameras. Examples of such systems range from affective music players (van der Zwaag et al. 2012) to intelligent car safety systems (Lisetti and Nasoz 2005) and air traffic control (Pantic et al. 2005).

Various studies have been performed using data obtained through video, using either still images, or a temporal sequence of images; with varying levels of success. Table 5.1 shows a brief summary of ten emotion classification studies from video. It can be seen that various parts of the body have been used to derive features from, that the number of features used varies heavily and performance ranges from 20% to 98% on various tasks. With exception of the 20% performance by Cottrell and Metcalfe (1991), the performances can be considered relatively high. Because the number of classes, number of participants, prior probability of classes and methods used for validation vary between these studies, their performance cannot be compared directly. In order to overcome this problem we choose a standard database for facial emotion recognition enabling us to benchmark our classifier performance.

As addressed by van den Broek et al. (2009), publicly available data sets that can be used as benchmarks are scarce in affective computing. For facial emotion recognition, however, such benchmark databases are available. Kanade, Cohn and Tian published a ″Comprehensive Database for Facial Expression Analysis″ in 2000 (Kanade et al. 2000), later known as the Cohn-Kanade database. It consists of image sequences displaying the faces of participants who were instructed to show a range of ″facial displays″ consisting of at least one Action Unit (AU). The participants were university students between 18 and 50 years of age, 69% of them female, and represented of a mix of ethnicities. The image sequences are labeled per still image with AUs that are active, which can be translated to emotion labels using a set of rules provided by Ekman et al. (2002). For 100 of the participants at least one of the prototypic emotions (Anger, Disgust, Fear, Joy, Sadness, and Surprise) has been recorded and can be used for the classification of emotions from facial expressions.

We observe that SVM is a very popular technique applied at this boundary of affective computing and computer vision. While SVMs have been applied successfully to various classification tasks, there are various reasons to investigate how alternative classification methods perform as affective classifiers. LVQ methods have been successfully applied in many settings (Neural Networks Research Centre, Helsinki 2002), but to the best of our knowledge, not to the task of recognizing facial expressions from the Cohn-Kanade database. This type of classifier has several benefits, such as low computational complexity resulting in fast training times, conceptually intuitive nature, and possibility to inspect relevant features without performing additional analyses. In order to put our work into perspective, we performed a comprehensive literature review of methods applied to the Cohn-Kanade database which will be treated in the next section. After that, we present the methods used for our affective classifiers and results obtained. Finally, we present a

Table 5.1: *Review of ten machine learning studies employing facial characteristics to recognize emotions.*

Reference	Input[1]	Ss[2]	Feat.[3]	Technique	Targets	Perf[4]
Cottrell & Metcalfe (1991)	Face [S]	20	4096	ANN	8 emotions	20%
Essa & Pentland (1995; 1997)	FACS [ST]	8			5 emotions	98%
Yacoob & Davis (1996)	Face [ST]	32	16		7 emotions	70%
Lien et al. (2000)	FACS [ST]	100	38	LDA, HMM	9 action units	80%
Cohen et al. (2003)	Motion Units [ST]	53	12	BN	7 emotions	83%
Pantic & Patras (2006)	FACS [ST]	19	24		27 action units	87%
Littlewort et al. (2006)	Face [S]	100	900	LDA, SVM	7 emotions	93%
Gunes & Piccardi (2009)	Head, hands, body [ST]	10	172	DT, BN, SVM, ANN, AdaBoost	12 emotions	85%
Sanchez et al. (2010)	Face [ST]	52	84	SVM	6 emotions	83%
Xiao et al. (2011)	Face [S]	$\leq$100	4320	kNN, SVM	6 emotions	97%

[1] Abbreviations used: S(patial), T(emporal)
[2] Number of subjects
[3] Number of features
[4] Performance (accuracy)

discussion and conclusion.

5.2 Cohn-Kanade database

The Cohn-Kanade database has been widely used to develop and validate techniques for facial emotion recognition. To obtain an overview of techniques and their performances, we have searched the literature systematically using Web of Science

Table 5.2: *Meta analysis of 150 models from literature.*

Nr. of Classes	Classes[1]	Validation method	Number of models	Accuracy min	mean	max
7	A,D,F,J,N,Sa,Su	cross-validation	31	78.90%	90.75%	99.40%
7	A,D,F,J,N,Sa,Su	pp-cross-validation	26	73.40%	85.94%	94.88%
6	A,D,F,J,Sa,Su	cross-validation	20	82.52%	89.21%	96.70%
6	A,D,F,J,Sa,Su	pp-cross-validation	17	76.12%	86.54%	96.40%
6	A,D,F,J,Sa,Su	single split	2	83.05%	87.43%	91.81%

[1] Abbreviations used: A(nger), D(isgust), F(ear), J(oy), (N)eutral, (Sa)dness, (Su)rprise

(Thomson Reuters 2014). The search terms ″Cohn AND Kanade″ resulted in 153 publications. We selected 43 publications for full analysis by excluding e.g., those using temporal information (video). In these papers we identified 199 classification schemes for 6 or 7 emotion classes, which were trained using the Cohn-Kanade database (Kanade et al. 2000) and for which a performance was reported.

We applied the following criteria for further filtering: accuracy of a model should be reported; and it should be validated using data from at least 50 participants, which is half of the available participants in the Cohn-Kanade database. These criteria were satisfied by 96 models, of which Table 5.2 shows a summary. First of all, it shows that the task of classifying unseen faces (using per person (pp)-cross-validation) is more difficult than classifying unseen instances of known faces (using cross-validation). Many studies concern the 6-class problem, considering the expressions Anger (A), Disgust (D), Fear (F), Joy (J), Sadness (Sa), and Surprise (Su). Alternatively, a 7-class problem which also includes a Neutral (N) expression has been investigated in a majority of studies. With exception of one study (Zavaschi et al. 2013) which reports exceptionally high performance of ensembles of SVMs: 99.40% accuracy, 3% more than the second best published result, the latter appears more difficult when judged by maximum performance. With respect to mean performances, however, this does not hold. This might be explained by the majority of research focussing on the 7-class problem.

Table 5.3: *Literature overview of studies that classify 7 emotion classes using the Cohn-Kanade database and validated using participant wise cross validation, grouped by feature type.*

Citation	Features	Classifier type	Accuracy	#pp used in validation	#images used in validation
Zhao and Zhang (2011)	KDIsomap	SVM	94.88%	96	1409
Zhao and Zhang (2011)	KIsomap	SVM	75.81%	96	1409
Zhao and Zhang (2011)	KLDA	SVM	93.32%	96	1409
Zhao and Zhang (2011)	LDA	SVM	90.18%	96	1409
Zhao and Zhang (2011)	KPCA	SVM	92.59%	96	1409
Zhao and Zhang (2011)	PCA	SVM	92.43%	96	1409
Jabid et al. (2010b)	LDP	SVM	93.40%	96	1632
Jabid et al. (2010b)	LDP	Template matching	86.90%	96	1632
Jabid et al. (2010b)	LBP	SVM	88.90%	96	1632
Shan et al. (2009)	LBP	SVM	88.90%	96	1280
Lajevardi and Hussain (2010)	LBP	LDA	88.40%	100	?
Zavaschi et al. (2013)	LBP	SVM	84.30%	100	1281
Shan et al. (2009)	LBP	Linear programming	82.30%	96	1280
Jabid et al. (2010b)	LBP	Template matching	79.10%	96	1632
Shan et al. (2009)	LBP	Template matching	79.10%	96	1280
Shan et al. (2009)	LBP	LDA	73.40%	96	1280
Shan et al. (2009)	LBP	LDA&ANN	73.40%	96	1280
Zavaschi et al. (2013)	LBP & Gabor	SVM ensemble	88.90%	100	1281
Zavaschi et al. (2013)	LBP & Gabor	SVM	79.20%	100	1281
Lajevardi and Hussain (2010)	HLACLF	LDA	91.60%	100	?
Lajevardi and Hussain (2010)	HLAC	LDA	89.90%	100	?
Lajevardi and Hussain (2010)	Gabor	LDA	89.70%	100	?
Jabid et al. (2010b)	Gabor	SVM	86.80%	96	1632
Shan et al. (2009)	Gabor	SVM	86.80%	96	1280
Lu et al. (2006)	Gabor	NKFDA	85.59%	93	≤ 651
Zavaschi et al. (2013)	Gabor	SVM	78.70%	100	1281

Table 5.4: *Literature overview of studies that classify 6 emotion classes using the Cohn-Kanade database and validated using participant wise cross validation, grouped by feature type.*

Citation	Features	Classifier type	Accuracy	#pp used in validation	#images used in validation
Jabid et al. (2010b)	LDP	SVM	96.40%	96	1224
Jabid et al. (2010b)	LDP	Template matching	89.20%	96	1224
Li et al. (2009)	SIFT	SVM	96.33%	90	300
Jabid et al. (2010b)	LBP	SVM	92.60%	96	1224
Shan et al. (2009)	LBP	SVM	92.60%	96	960
Shan et al. (2009)	LBP	Linear programming	89.60%	96	960
Jabid et al. (2010b)	LBP	Template matching	84.50%	96	1224
Shan et al. (2009)	LBP	Template matching	84.50%	96	960
Shan et al. (2009)	LBP	LDA	79.20%	96	960
Shan et al. (2009)	LBP	LDA&ANN	79.20%	96	960
Jabid et al. (2010b)	Gabor	SVM	89.80%	96	1224
Shan et al. (2009)	Gabor	SVM	89.80%	96	960
Fazli et al. (2009)	Gabor & PCA&LDA	PNN	89.00%	70	192
Wang and Yin (2007)	facial feature points	LDA	82.68%	53	864
Wang and Yin (2007)	facial feature points	QDC	81.96%	53	864
Wang and Yin (2007)	facial feature points	SVC	77.68%	53	864
Wang and Yin (2007)	facial feature points	NBN	76.12%	53	864

We focus on the most difficult cross validation type (pp-cross-validation) that assesses the performance on classifying emotions from unseen faces. Table 5.3 shows the 26 classifiers that are published for the 7-class problem, showing that accuracy ranges between 73.40% (Shan et al. 2009) and 94.88% (Zhao and Zhang 2011). The most used feature-type is LBP followed by Gabor features. Highest performance is obtained by methods that use (non-)linear projections of the original images to some lower dimensional space, such as KDIsomap, LDA and PCA. Slightly inferior are the methods based upon feature extraction such as LDP and LBP and less successful are methods based upon Gabor features. The most popular classification techniques are SVM and LDA.

Table 5.4 shows the 16 classifiers found for the 6-class problem showing performances ranging from 76.12% (Wang and Yin 2007) to 96.40% (Jabid et al. 2010b). Similar trends as for the 7-class problem can be observed where LBP features are most

used. LDP features reach highest performance closely followed by Scale-Invariant Feature Transform (SIFT) features (Li et al. 2009) and LBP. Again, Gabor features perform worse, followed by facial features points; and SVM dominates among the types of classifier used.

In this work, we will explore the application of LVQ classifiers to this classification problem and will use the open box nature of these classifiers to gain more insight into the classification problem. We will use LBP-features because they have been used most often, can be obtained relatively efficiently and have been demonstrated to give good performance.

Figure 5.1: *Examples of cropped images (Kanade et al. 2000) representing (from left to right) Anger, Disgust, Fear, Joy, Sadness and Surprise.*

5.3 Methods

From the Cohn-Kanade database we selected 310 image sequences, coming from 95 subjects, that could be labeled as one of the emotions Anger, Disgust, Fear, Joy, Sadness or Surprise. For each sequence, the neutral face and three peak frames, i.e., those with highest emotional intensity, were used for emotional expression recognition. Following Shan et al. (Shan et al. 2009, Gritti et al. 2008) and Tian (2004), we used the distance between manually annotated location of the eyes to rotate, crop and scale the images to 108x147 pixels, which were used as input to the further preprocessing. First, the images were rotated to ensure horizontal alignment of the eyes. The distance between the eyes (d_{eyes}) was determined and then the images were cropped such that they measured $2d_{\text{eyes}}$ by $3d_{\text{eyes}}$, and finally they were resized to 108x147 pixels. Figure 5.1 shows examples of resized images from several participants. As discussed in the previous section, the most used feature type is LBP, which is the one of our choice. We derived the LBP-features from the scaled images in the following way:

Per grey valued pixel (i_c) the LBP-value is calculated by comparing the pixel to its eight neighbors, resulting in a binary string of which the decimal value is taken, according to:

$$LBP(x_c, y_c) = \sum_{n=0}^{7} s(i_n - i_c)\, 2^n \tag{5.1}$$

where s is the Heaviside step function. The $2^8 - 1$ possible outcomes are reduced to $L = 59$ by regarding only those LBP values with at most 2 bitwise transitions when considered as a circular pattern, as proposed by Ojala et al. (2002). The patterns in this subset are termed "Uniform" patterns and represent bright and dark spots, corners and edges.

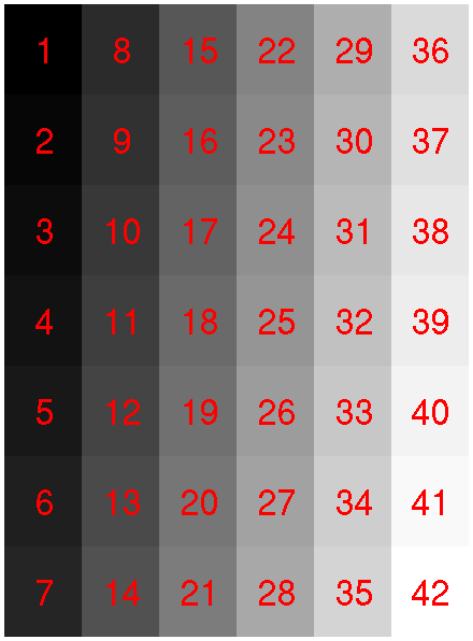

Figure 5.2: *Subdivision of an image into 42 regions. The numbers indicate the ordering used in the concatenation of the feature vectors.*

The images are divided into $(6 * 7 = 42)$ regions R_j of size 18x21 pixels, as indicated by Figure 5.2, where per region a histogram $H_i = \sum_{x,y} \delta_{LBP(x,y),i}$ (with $(x, y) \in R_j, i = 0, \ldots, L - 1$) is built. These histograms are placed next to each other, forming a single vector of length $N = 42 * 59 = 2478$. Another variant uses

$11*13 = 143$ overlapping regions, resulting in a vector of length $N = 143*59 = 8437$.

Validation was performed using 10x10-fold participant-wise cross validation, i.e., 10 repetitions of randomly chosen 10-fold cross validation, where participants are strictly separated in training and test data. In this way, the performances obtained reflect the generalization performances to unseen participants. We applied the classifiers to both 6- and 7-class facial expression recognition tasks, compared their generalization performances, and inspected the confusion matrices. Finally, we inspected the prototypes trained by Robust Soft Learning Vector Quantization (RSLVQ), with specific attention for the relevances it (implicitly) assigns to the features. To this end, we inspected differences between the 'Neutral' prototype and other prototypes.

5.4 Results

The results, given in Table 5.5 and 5.6 show that for the subtask of classifying 7 emotion classes, our classifiers reach over 91% accuracy. RSLVQ competes well with SVM, the latter reaching slightly higher performance, especially when using the LBP features with overlapping regions. On the other hand, the results over the 10 times 10-folds are slightly more stable for RSLVQ. For the 6-class classification task, we observe a similar pattern: very good performance of RSLVQ (93.3%) and even slightly better accuracy by SVM (up to 94.5%). Tables 5.7 and 5.8 report the training performances obtained for the 7-class and 6-class tasks, respectively. We observe that both SVM and RSLVQ manage to achieve perfect training performance.

Table 5.5: *Test performances on* 10×10 *fold participant-wise cross validation on the 7-class facial expression data.*

	LBP overlap		LBP no overlap	
Method	Hyper-parameter	Accuracy	Hyper-parameter	Accuracy
Baseline		21.6%		21.6%
kNN	$k = 11$	$73.8\% \pm 0.7\%$	$k = 11$	$72.2\% \pm 0.7\%$
SVM - linear	$C = 0.1$	$92.2\% \pm 0.5\%$	$C = 0.1$	$91.4\% \pm 0.5\%$
Means		$83.1\% \pm 0.6\%$		$82.8\% \pm 0.4\%$
GLVQ		$84.0\% \pm 0.3\%$		$82.9\% \pm 0.4\%$
RSLVQ	$v_{soft} = 5$	$91.3\% \pm 0.3\%$	$v_{soft} = 0.6$	$91.2\% \pm 0.5\%$

Table 5.6: *Test performances on* 10×10 *fold participant-wise cross validation on the 6-class facial expression data.*

	LBP overlap		LBP no overlap	
Method	Hyper-parameter	Accuracy	Hyper-parameter	Accuracy
Baseline		26.2%		26.2%
kNN	$k = 11$	$79.6\% \pm 0.7\%$	$k = 11$	$78.7\% \pm 0.7\%$
SVM - linear	$C = 0.01$	$94.5\% \pm 0.6\%$	$C = 0.1$	$94.0\% \pm 0.4\%$
Means		$88.5\% \pm 0.4\%$		$87.2\% \pm 0.3\%$
GLVQ		$85.4\% \pm 0.4\%$		$84.5\% \pm 0.3\%$
RSLVQ	$v_{soft} = 5$	$93.2\% \pm 0.5\%$	$v_{soft} = 0.9$	$93.3\% \pm 0.2\%$

Table 5.7: *Training performances on* 10×10 *fold participant-wise cross validation on the 7-class facial expression data.*

	LBP overlap		LBP no overlap	
Method	Hyper-parameter	Accuracy	Hyper-parameter	Accuracy
Baseline		21.6%		21.6%
kNN	$k = 11$	$87.7\% \pm 0.0\%$	$k = 11$	$88.0\% \pm 0.0\%$
SVM - linear	$C = 0.1$	$100.0\% \pm 0.0\%$	$C = 0.1$	$100.0\% \pm 0.0\%$
Means		$90.9\% \pm 0.1\%$		$90.3\% \pm 0.1\%$
GLVQ		$93.9\% \pm 0.1\%$		$93.3\% \pm 0.1\%$
RSLVQ	$v_{soft} = 5$	$100.0\% \pm 0.0\%$	$v_{soft} = 0.6$	$100.0\% \pm 0.0\%$

Table 5.8: *Training performances on* 10×10 *fold participant-wise cross validation on the 6-class facial expression data.*

	LBP overlap		LBP no overlap	
Method	Hyper-parameter	Accuracy	Hyper-parameter	Accuracy
Baseline		26.2%		26.2%
kNN	$k = 11$	$87.3\% \pm 0.1\%$	$k = 11$	$87.9\% \pm 0.0\%$
SVM - linear	$C = 0.01$	$100.0\% \pm 0.0\%$	$C = 0.1$	$100.0\% \pm 0.0\%$
Means		$93.7\% \pm 0.1\%$		$93.1\% \pm 0.1\%$
GLVQ		$93.5\% \pm 0.1\%$		$93.6\% \pm 0.0\%$
RSLVQ	$v_{soft} = 5$	$100.0\% \pm 0.0\%$	$v_{soft} = 0.9$	$100.0\% \pm 0.0\%$

Confusion matrices of SVM and RSLVQ are available in Tables 5.9 and 5.10. Differences between the confusions made by both classifiers are small and for both we observe that most errors correspond to misclassifying various emotions as 'Neutral'. This might suggest that the classifiers have most difficulty with low-intensity instances of emotions (other than Neutral) while the emotions themselves are quite well separable. Most difficult emotions are Fear, of which 13% is misclassified as Joy, and Anger, which is often mistaken with Neutral and Sadness. We also inspected the confusion matrices for 6-class classification (see Tables 5.11 and 5.12) and noticed that the misclassifications as Neutral are mainly compensated by increased class-wise accuracy of Sadness and slight increases for the other emotions, of which Surprise can be detected flawlessly.

In order to inspect the (implicit) relevances assigned by RSLVQ, we summed up

Table 5.9: *Confusion matrix (averaged over 10x10-fold cross validation) for 7class classification by SVM. Entries are percentages per actual emotion.*

Actual \ Predicted	A	D	F	J	N	Sa	Su
Anger	78.8	3.3	0	0.1	10.7	7.1	0
Disgust	3.6	90.3	0	0	2.3	3.7	0
Fear	0	0	78.6	12.6	6.8	1.9	0.1
Joy	0	0	0.2	99	0.9	0	0
Neutral	0.5	0	0	1	94.5	2.7	1.3
Sadness	2.6	0.3	0	0	9	85.6	2.5
Surprise	0	0	0	0	1	0	99

Table 5.10: *Confusion matrix (averaged over 10x10-fold cross validation) for 7class classification by RSLVQ. Entries are percentages per actual emotion.*

Actual \ Predicted	A	D	F	J	N	Sa	Su
Anger	81.8	5.9	0	0.6	7.9	3.7	0
Disgust	2.1	93.3	0	0	1.8	2.8	0
Fear	1.2	0.4	79.4	13.5	2.3	2.2	1
Joy	0.4	0	0.7	97.7	1.3	0	0
Neutral	1	0.2	0	1.5	93.2	2.4	1.8
Sadness	3.9	0.8	0	0.3	7.8	84	3.2
Surprise	0	0	0	0	1.4	0	98.6

all absolute pairwise differences between the prototype representing 'Neutral' faces and the other emotions. The difference vector of two prototypes corresponds to the direction in feature space along which the two classes are discriminated. The absolute value of its components can be interpreted as to measure the relevance of the corresponding feature. Figure 5.3 shows this information aggregated per region (as used in the building of the LBP histograms). It indicates that most informative to the classifier are the regions around the mouth, followed by the eyes and eye-brows.

The feature vectors we used represent the frequency of observing certain textural elements within 42 different regions of the face. Figure 5.4 shows the LBP-features linked to the 48 most relevant histogram entries. We see that, out of the 42 regions, the regions around the mouth are best represented. Regions 20 and 27 represent the upper side of the mouth and the LBP-features represented in the top 48 indicate the importance of textural components that are lighter at the top than on the bottom. Similarly, regions 21 and 28 represent the lower side of the mouth, from which LBP-

Table 5.11: *Confusion matrix (averaged over 10x10-fold cross validation) for 6class classification by SVM. Entries are percentages per actual emotion.*

Actual \ Predicted	A	D	F	J	Sa	Su
Anger	83.7	3.2	0	2.5	10.6	0
Disgust	3.1	93.3	0	0	3.6	0
Fear	0	0	83.5	14	2.3	0.1
Joy	0.3	0	0.2	99.4	0.2	0
Sadness	3.4	0.4	0	0.1	94.1	2
Surprise	0	0	0	0	0	100

Table 5.12: *Confusion matrix (averaged over 10x10-fold cross validation) for 6class classification by RSLVQ. Entries are percentages per actual emotion.*

Actual \ Predicted	A	D	F	J	Sa	Su
Anger	82.5	6	0	2.4	9.1	0.1
Disgust	1	96.4	0	0	2.6	0
Fear	0.3	0	79.9	15.4	2.6	1.8
Joy	0.4	0	0.6	99	0	0
Sadness	3	1	0	0.1	91.8	4.1
Surprise	0	0	0	0	0	100

features that indicate lighter bottom and darker top are present. Finally, we observe that regions 13 and 34, corresponding to the left and right side of the mouth, are mostly represented by textural components that have darker right and left sides, respectively. These observations seem to indicate that opening of the mouth, which is accompanied by dark pixels in the center of the mouth-region, is the most important distinction between various emotions.

Figure 5.5 shows the aggregated relevances per region for each of the emotions in isolation, i.e., representing the difference to the 'Neutral' emotion. We observe that the relevances for Surprise are quite distributed, and more expressive around the central mouth regions and chin. Sadness shows even higher relevance of the chin areas; Joy is most different from Neutal in the outer and upper mouth regions, while Fear differs in the outer and central mouth regions. Finally, the relevances of Anger and Disgust are more scattered, but in comparison to the other emotions have relatively high contributions of the features from the eyes, brows and forehead.

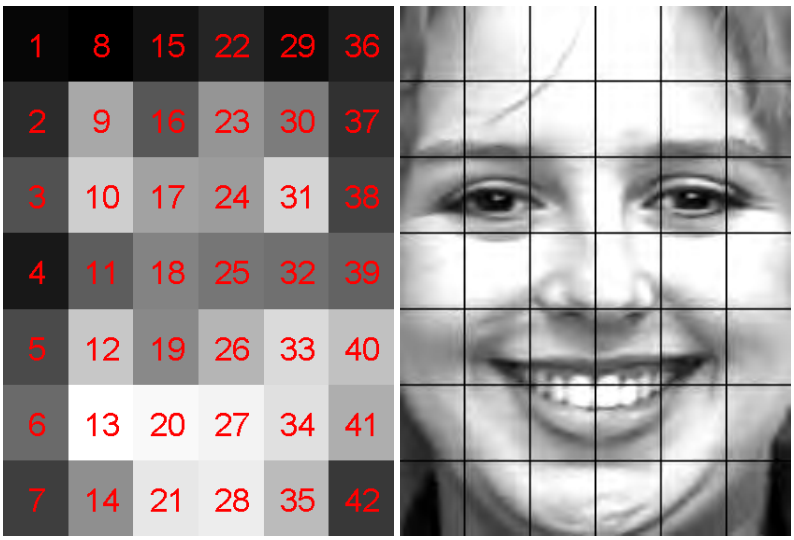

Figure 5.3: *Relevance of image regions in the RSLVQ classifier (left;white levels indicate relevance), and example picture for reference (right).*

5.5 Discussion

The results we obtained show high accuracy on the tasks of classifying facial expressions, represented as LBP feature vectors, into emotions. We used two different representations, one using non-overlapping regions in the facial pictures, the other using overlapping regions. The non-overlapping regions yield more intuitively interpretable feature vectors, while the overlapping regions contain more information, but also increase the dimensionality of the feature vectors almost by a factor 4. Using non-overlapping regions, the 6-class classification task was best performed by SVM with 94.0% accuracy. Second best was RSLVQ with 93.2%. Slightly better performances were obtained using the overlapping regions (maximum accuracy of 94.5%). When comparing these to the performances reported in literature on the same data set and classification task, we observe that there are two attempts that gained better performance. Both use SVM and either use LDP (Jabid et al. 2010b) or SIFT features (Li et al. 2009). The SIFT-based study, however uses only a subset of 300 pictures, indicating that they left out pictures, which might be suitably chosen, rather than the full data set. If we compare our method with other methods using LBP features, we outperform them by 1.9 percentage points.

For the 7-class classification task, which includes the neutral face as a class, our

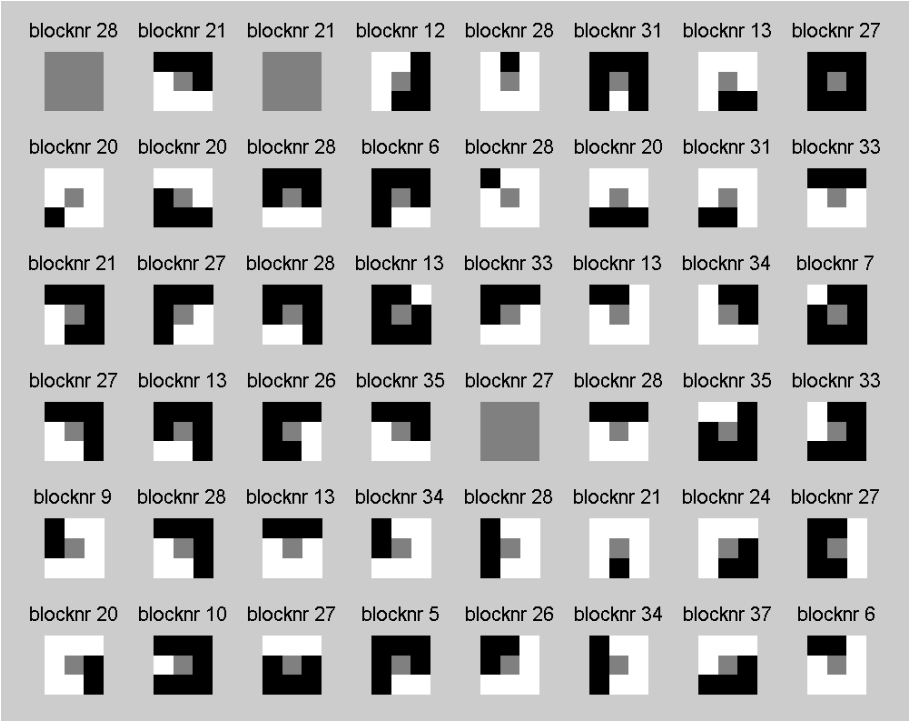

Figure 5.4: *Top 48 most relevant LBP-features used the RSLVQ classifier; ordered from let to right, top to bottom. The block numbers refer to the regions as numbered in Figure 5.3.*

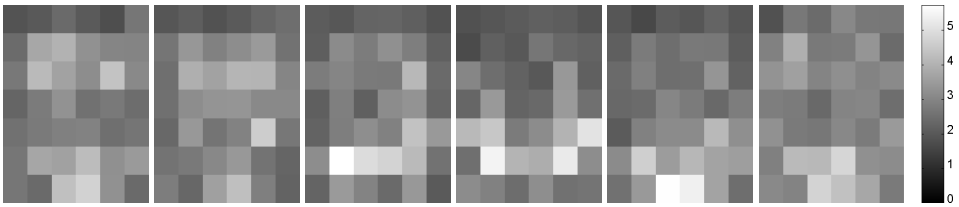

Figure 5.5: *Relevance of image regions in the RSLVQ classifier for Anger, Disgust, Fear, Joy, Sadness and Surprise (from left to right).*

classifiers reached an accuracy of 91.4% for SVM and 90.9% for RSLVQ. Again, slightly better performances were obtained using the overlapping regions (maximum accuracy of 92.2%). Four techniques (Zhao and Zhang 2011) from one paper show better performances when using SVM and different feature sets. In comparison to the methods that use LBP features, our classifiers perform better by 3.3

percentage points.

The prototype based classifiers we used enabled us to inspect the prototypes and infer which features are considered most influential by the classifiers. The mouth region turned out to be most influential. Within this region the LBP features that correspond to various mouth openings were most important. While eye-brows are known to be activated in many different emotions (Ekman 1979), and they are found to be the prominent facial elements to highlight prosody (Swerts and Krahmer 2008), our results suggest that for automated facial expression recognition, the mouth-region is more important. The regions representing the eye-brows and forehead, however do help our classifier in distinguishing especially Anger and Disgust from the other emotions. Shan and colleagues (Shan et al. 2009) used AdaBoost in combination with pattern matching to determine the most influential LBP histograms from an exhaustive set of 16640 facial regions and identified most discriminant regions around the eyes and mouth. With our approach, we obtained these indications of relevance directly from the trained classifier, rather than performing additional and computationally intensive analyses.

We also observed that not only the occurrence frequencies of uniform LBP features are relevant for the classification, but also the frequency of non-uniform patterns, which were joined together in one bin in each histogram representing a photo region, were represented in the list of most influential features. Moreover, the 1st, 3rd and 29th most influential features were such non-uniform patterns. On the other hand, the use of uniform patterns rather than all LBPs reduced the feature space with more than a factor 4 and helps keeping the search space manageable.

5.6 Conclusion

We have performed a comprehensive literature overview of attempts to classify facial expressions from the Cohn-Kanade database and observed that generalization performances on the 7 and 6 class tasks average at 85.9% and 86.5%, respectively. Maximum reported accuracies on these tasks were 94.9% and 96.4%. While being the most popular, or at least most frequently used, type of features, LBP features reached only up to 88.9% and 92.6%, respectively for 7 and 6 classes.

To the best of our knowledge, we have applied LVQ classifiers for the first time to the task of facial expression recognition using the Cohn-Kanade database. The generalization accuracies obtained (91.3% for 7-class, and 93.3% for 6-class classifi-

cation) show that RSLVQ is among the most successful classifiers overall and outperforms all reported efforts using LBP features. As a reference we used SVM, which showed even slightly better performances (92.2% for 7-class, and 94.5% for 6-class classification) but, in contrast to RSLVQ, does not allow for direct inspection of the knowledge learned and used by the classifiers. By inspecting the prototypes trained by RSLVQ we noticed that the most prominent features originate from the mouth region, followed by the eye-regions. The specific LBP features that are used most prominently by the classifier confirm that mouth opening/closing is discriminative for various emotions.

In the present work, we have used implicit relevances obtained from difference vectors of RSLVQ prototypes. Other LVQ variants can be designed that explicitly train relevance vectors along with prototypes; examples are Generalized Matrix Learning Vector Quantization (GMLVQ) (Schneider et al. 2009a) and Matrix Robust Soft Learning Vector Quantization (MRSLVQ) (Schneider et al. 2009b). Future work includes the application of such methods to the challenge of facial expression recognition. Another interesting future extension of the current work is to observe how our methods perform on spontaneous emotions. Although being more challenging, recent developments (Wan and Aggarwal 2014) indicate that results obtained in one setting can be transferred successfully to the other. Finally, the literature review we performed indicates that performances might be further improved by considering different feature sets such as LDP or SIFT.

The high performances obtained indicate that implementation in consumer products starts to become feasible. Natural choices of first applications include real time behavior adaptation of laptops or tablets to their users' emotions. By, for example, being able to distinguish frustration from happiness, human-computer interaction can be greatly improved because it allows for detection of suboptimal interactions and adapt at real time by offering alternative actions when frustration is detected. The limited complexity of LVQ, that can directly handle multi-class classification (i.e., without requiring classification schemes such as 'one-vs-all' that are needed by binary classifiers), allows for quick training times and opens up the ability to train user specifics and personalize the model by learning at real time. Such personalized systems should be able to obtain even better performances for facial expression recognition.

Material based on:
de Vries, J. J. G., Lemmens, P. M., Brokken, D., Pauws, S. C. and Biehl, M.: in press, Towards Emotion classification using appraisal modeling, International Journal of Synthetic Emotions.

Hoofdstuk 6

EMOTION FROM A COGNITIVE PERSPECTIVE

Abstract

We studied whether a two-step approach based on appraisal modeling could help in improving performance of emotion classification from sensor data that is typically executed in a one-stage approach in which sensor data is directly classified into a (discrete) emotion label. We propose a two-stage method in which sensor data are first represented in terms of appraisal dimensions and only then classified to emotion labels. The proposed intermediate step is inspired by appraisal models in which emotions are characterized using appraisal dimensions (Scherer 1993), and subdivides the task in a person-dependent and person-independent stage.

A two-stage approach has two main benefits: it can use dedicated techniques to deal with the specifics of both stages (e.g., person-dependency) and by design, appraisal models should work in all affective contexts, enabling re-use of the second-stage classifiers in other affective applications.

In this paper we assessed feasibility of this second stage: the classification of emotion from appraisal data. To this end, we gathered data from 134 participants that empathized with emotional events visualized on photos, and scored both appraisal dimensions and emotion labels for the emotions that they experienced. We applied a variety of machine learning techniques and used visualization techniques to gain further insight into the classification task. Appraisal theory assumes the second step to be independent of the individual. Results obtained are promising, but do indicate that not all emotions can be equally well classified, perhaps indicating that the second stage is not as person-independent as proposed in the literature.

6.1 Introduction

One type of challenges within the domain of affective computing are various classification challenges like for instance affective annotation of images, video, or written text; or the recognition of various affective states given sensor data. In this work we focus at the subdomain of emotion classification from sensor data, which can be based on many different input signals. The field is dominated by

three main input types, all measuring signals from the human body: video, audio and physiology. Emotional states can be derived from video through facial expressions, postures or movements (Gunes and Piccardi 2009, Sanchez et al. 2010, Xiao et al. 2011, van Kuilenburg et al. 2008); from audio through utterances (Sobol-Shikler and Robinson 2010, van den Broek et al. 2011, Wu et al. 2011); and from physiology through a variety of bodily signals such as cardiac activity, skin conductance and respiration (Chanel et al. 2009, Hosseini et al. 2010, van den Broek, Lisý, Janssen, Westerink, Schut and Tuinenbreijer 2010). We refer to Janssen, Tacken, de Vries, van den Broek, Westerink, Haselager and IJsselsteijn (2013) for a more comprehensive overview of studies that use one or more of these input signals for emotion classification.

Despite various advances in the field, performances are generally below those in other fields of automated classification, such as finger print recognition, restricted cases of handwriting recognition, etc. This indicates that affective classification tasks are generally difficult. We also observe that most automated affective recognition systems use a one-step approach, directly mapping measured features to emotion labels. As suggested by some (Ptaszynski et al. 2009), the integration of contextual information in a multi-step approach might aid in the interpretation of various input signals. Because this approach might be applicable to a wide range of different affect recognition systems it calls for a structural approach.

We hypothesize that models coming from appraisal theory might provide such a reasoning framework in which factual contextual information can be combined with other sensor data that carry information about the personal interpretation of the person being measured. We propose a system that uses an appraisal model in a two-step approach, in which the first step maps measurement data to an appraisal representation, and a second step maps the appraisal representation onto emotion labels. Several appraisal theories have been coined (Ortony et al. 1988, Scherer 2001, Frijda 1987, Lazarus 1991, Marsella et al. 2010) that have in common that they propose the process towards appraisal of stimuli to be person dependent whereas the generation of emotion from an appraisal is proposed to be person-independent. This means that interpersonal differences can be taken into account in the first step of the suggested two-step approach, while the second step is independent of the user of the system.

Regarding the mapping of sensor data to appraisal, several studies have shown that various appraisal dimensions can be obtained from, e.g., physiological measurements (Aue et al. 2007, Grandjean and Scherer 2008, Bradley et al. 1993, Smith

1989, van Reekum et al. 2004). The second step of mapping appraisals to emotion labels, has been studied to certain extent by the authors of appraisal models (Scherer 1993, Scherer et al. 2006), but apart from one recent publication (Meuleman and Scherer 2013) did not involve more sophisticated classification techniques from machine learning research. Our aim is to research the potential use of appraisal models in a two-step approach to emotion classification. In the present study we will provide an independent assessment of one such appraisal model using a variety of machine learning techniques and use visualization techniques to gain further insight into the classification task.

Given the limited performance observed in existing affective classification systems, the benefit of the two-step approach we propose is that it will separate the person-dependent and person-independent variability in the data. In the two separate steps, dedicated classification techniques can be applied that specifically address the different sources of variation. Thereby the overall system could come to a better performance than the one-size-fits-all approach of one-step classification can achieve. In addition, from a research perspective, the two-step approach enables separate study of both steps while still enabling easy integration (presuming a common framework).

In the next section, we will continue with an overview of appraisal theory. Section 6.3 sketches the field of emotion classification and the potential of appraisal theory in this respect. In Section 6.4 we will discuss the experiment performed, of which the results are presented in Section 6.5. Finally we end with a discussion, conclusion and outlook in Sections 6.6 and 6.7.

6.2 Appraisal Theory

An appraisal (sometimes also called a construal) is an interpretation of the causes and consequences of a stimulus given the perceiver's personal goals, standards, and attitudes (Ortony et al. 1988). Because of the involvement of personalized cognitions like goals, such stimuli are often interpreted as being beneficial or a hindrance in achieving a goal. Thus, the element of valence, a positive or negative connotation of a stimulus, is introduced. This element is what distinguishes an emotional from a non-emotional response that both can result from observing and mentally processing a stimulus. An elaborate overview of computational models of emotions, including those based on appraisal theory is provided by Marsella et al. (2010). Currently, two theories dominate the field: Scherer's Component Process Model

(CPM) (Scherer 2001) versus the OCC model of Ortony et al. (1988). In addition to the CPM and OCC model, the theories of Frijda (1987) and Lazarus (1991) theories complete the palette of appraisal theories. The OCC model is particularly focused on describing the concepts and their inter-relations that are needed to explain how the construal of stimuli results in an emotion. We chose to use Scherer's CPM to build our predictive computational model of the appraisal process because it describes the relation between stimulus and emotion in more detail and on different levels of granularity. A more recent elaboration by Grandjean and Scherer (2008) also links the cognitive states of the CPM to neuronal systems.

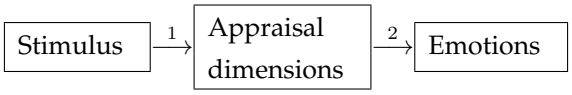

Figure 6.1: *Appraisal model represented as a two-stage process.*

Similar to Scherer (1993), we reduced the complex CPM to a two-stage strategy to cover the steps that are between the perception of a stimulus and the (overt) display of an emotion (see Figure 6.1). This simplification obviously removes many elements that are characteristic for the CPM but enables us to formulate a computational model that can be validated using collected data. Furthermore, the two-stage strategy applied in the current study is in line with the main computational appraisal architecture presented in Marsella et al. (2010), except for the feedback loop that enables complex emotions to build up from sequences of (other) emotional reactions. Because we are more interested in single emotional responses instead of the complex interaction between sequential emotions, this simplification is not limiting our research goals.

The first stage of the process builds up a representation of the stimulus in terms of the appraisal dimensions, as specified by the CPM model. This stage results in the specification of what we call an appraisal vector which is a collection of values for each of the appraisal dimensions. This vector is used in the second stage to map the appraisal dimensions onto an emotion. The experiments described in this paper focus on the second stage. That is, given an appraisal, to determine the corresponding emotion.

Marsella et al. (2010) also consider a two stage approach but takes as input the 'person-environment-relationship', rather than a 'stimulus'. One might argue that

a 'stimulus' is only part of the 'person-environment-relationship'. However, following the theory of appraisal, the personal interpretation based upon the context will be part of the first stage. This implies that the person-environment-relationship will be reflected in the appraisal vectors. Through this mechanism the other aspects of the 'person-environment-relationship' are represented and taken into account in our approach.

For the experiments, we used the full set of 15 appraisal dimensions: Novelty, Time, Intrinsic Pleasantness, Relevance, Expectedness, Conduciveness, Urgency, Causality by self, Causality by other, Causality by Chance, Control, Power, Adjustment, External Normative Significance and Intrinsic Normative Significance. The emotions covered by the appraisal model are Enjoyment, Joy, Disgust, Contempt, Sadness, Desperation, Anxiety, Fear, Irritation, Anger, Indifference, Shame, Guilt, and Pride (Scherer 1993).

6.3 Emotion Classification

Over the years, many different approaches have been tried to classify collected data that describes the current state of a user into emotion categories. These include a wide variety of measuring modalities, ranging from physiological parameters to audio and video captured from users. Janssen, IJsselsteijn, Westerink, Tacken and de Vries (2013) provides a literature overview on the classification of emotions using different measurement modalities and shows that performances between 50 and 70% are quite common, while also reports of 21% and, on the other extreme, 98% are available. To fully interpret and compare the reported performances, a more thorough study would be needed that also takes into account the number of classes used, circumstances during which the data was collected and protocols used for classifier validation. This goes beyond the scope of this paper; nevertheless, the observed performances clearly indicate that emotion recognition as classification task has not been resolved yet and that there is room for improvement.

We investigated whether the framework of appraisal models provided a useful divide-and-conquer approach for emotion recognition. In such an approach the classification task is turned into a two-stage classification in which first a mapping is made of a measurement of the stimulus to an appraisal vector, and from that, a second classification transforms the appraisal vector into a discrete emotion label. We specifically studied the second step not because the first step is trivial but because we think that it is in this step that significant performance gains can be achieved.

In itself, the required performance for an affective classifier is difficult to establish as it heavily depends on the application. Let us therefore consider accuracy of humans judging emotional cues. To our knowledge, there is no experiment that shares appraisal data to participants with the task to classify that into emotions. This might be due to the latent, non-observable character of appraisal. In another experiment (Janssen, Tacken, de Vries, van den Broek, Westerink, Haselager and IJsselsteijn 2013) participants had to categorize another person describing an emotional experience into one of 5 emotional classes (Happy, Relaxed, Sad, Angry, and Neutral). They had access to video and audio material (but the ability to judge from semantic information was taken out by only including participants that did not understand the language spoken by the participants in the video and audio material). Janssen and colleagues found that humans reach an accuracy of up to 31.0%, and less when only one modality was provided (22.7%, and 26.1% for audio- and video-only, respectively). The 5 classes were fully balanced; therefore random guessing would lead to a performance of 20%, thus humans performed hardly above random guessing when provided with information from a single modality. It is safe to assume that any computerized system should at least perform on par with humans in order to be accepted. While it is difficult to extrapolate these performances to the classification task involving appraisal data because we would compare full (human) emotion recognition from signals versus step two of a two-stage automated classification involving appraisal data, we believe it is safe to assume that the human performance could indicate a lower bound on the performance we should aim at for the complete two-step emotion classification.

6.4 Method

We carried out two experiments to gather sufficient data to be able to apply machine-learning techniques. Most of the data was gathered via a web-based experiment, but an experiment in the lab served as a precursor for developing the web-based version.

6.4.1 Lab Experiment

Participants

In the lab experiment, data were collected from 33 voluntary participants. The data of one participant was left out because this participant indicated afterwards that he did not understand all questions. Therefore the remaining data set contained data

from 32 participants (18 were male; mean age 27 years). Each participant rated 5 pictures, yielding 160 appraisals of which one instance was discarded because all questions were answered with ″not applicable″ (yielding 159 appraisals in total).

Stimuli

For each of the 14 emotions covered by the appraisal model we selected two pictures resulting in 28 pictures used. Twelve of the pictures (marked in Figure 6.2) originate from a validated picture set of Overbeek et al. (2007) and targeted the emotions of Disgust, Enjoyment, Fear, Irritation, Joy and Sadness. Because not all emotions covered by the appraisal model were present in this picture set, we added pictures found on the web, targeting at the remaining emotions of Anger, Anxiety, Contempt, Desperation, Indifference, Pride, Shame, and Guilt. Out of this set of 28 photos, depicted in Figure 6.2, for each subject a number of photos were chosen randomly (with the restriction that no photo should be chosen more than once per subject) in such a way that over all participants all photos were presented an approximately equal number of times. The number of photos to be rated per participant was chosen to be five.

Procedure

The experiment started with general demographic questions (i.e., gender and age). Subsequently, one of the five randomly selected photos was shown full screen for 6 seconds. For each photo, the participants were asked to imagine being in the event during which the photo was taken, to assume a role in that event, and to fill in an 18-item questionnaire. Questions 1 and 2 probed for the most prominent emotion triggered by a photo and how strong it was experienced. This emotion label will be used in the classification task as ground truth. The third question was to give a description of the assumed role of the participant and the situation s/he imagined. This question was added to help a participant stick to a single role while filling in the remaining 15 items of the questionnaire. These fifteen questions, that measure the fifteen appraisal dimensions, were taken from Scherer (1993), and adapted to fit the photo experiment (rather than recollection). They are listed in Table 6.1, and used the same 6-item Likert scales as the original. We only changed original labels ″not pertinent″ to ″not applicable″ because we found out in a pilot that many participants were familiar with the phrasing. Note that while all questions use unipolar or bipolar scales for the answer, the phrasing of the questions suggest a binary (yes/no) answer. Although we consider this questionnaire suboptimal, we stayed as close as possible to the original to be able to compare the results.

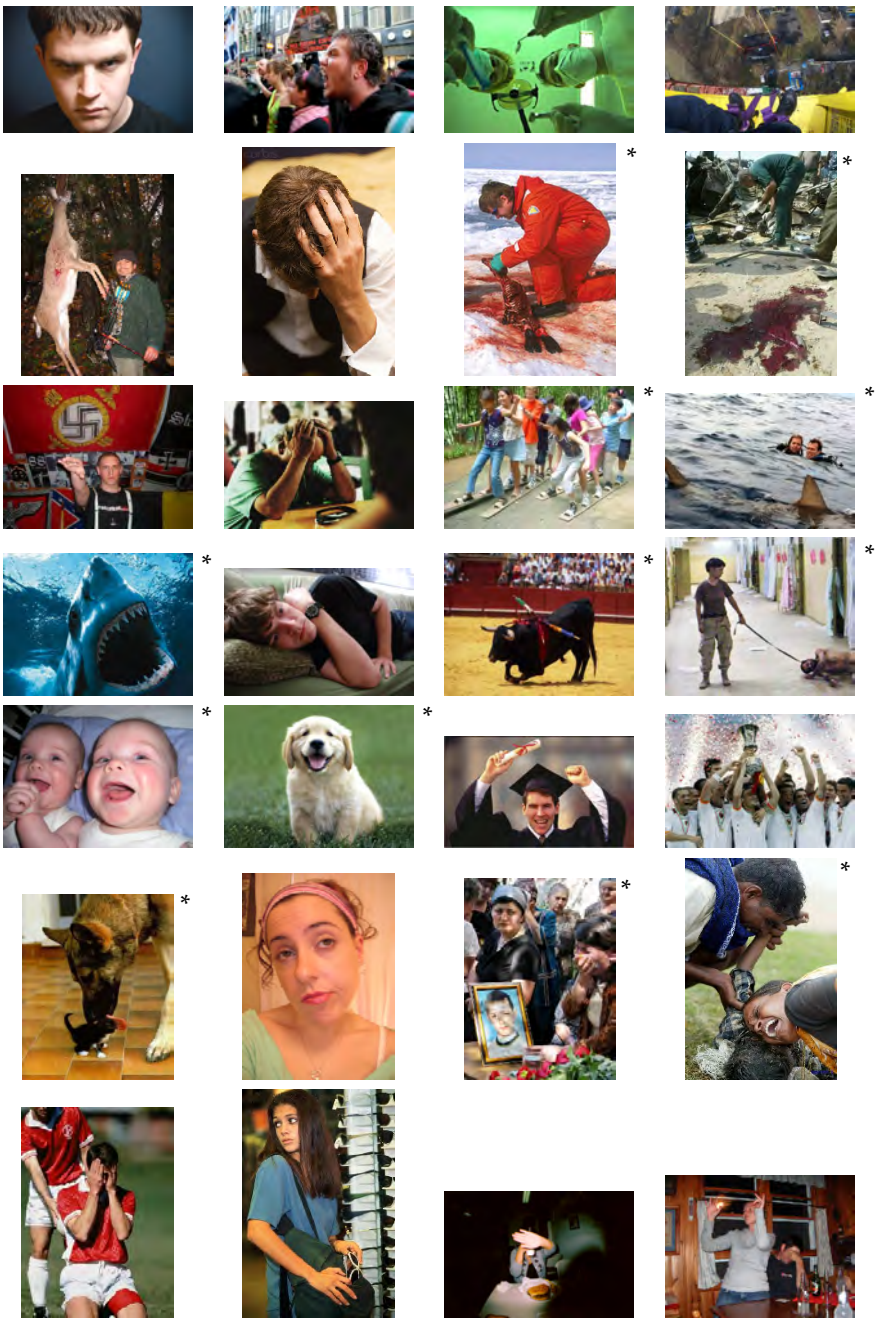

Figure 6.2: *Thumbnails of the pictures used for emotion induction. Pictures marked with an asterisk (* on the right top), have been taken from Overbeek et al. (2007).*

During the questionnaires a thumbnail of the corresponding photo was shown in the upper left corner of the screen, again to help participants stick to a single interpretation throughout the questionnaire. We used the answers to the first three questions to study the range of interpretations of (a particular) photo(s) and to verify that answers on the appraisal questions that followed could be expected. Note that, by following this procedure, we fully isolated the process of obtaining ground truth from the validation of the classifier performance to avoid biased performance estimation.

Table 6.1: *Questions used to measure the fifteen appraisal dimensions, adapted from (Scherer 1993).*

Appraisal dimension	Corresponding question
Novelty	Did the situation showed by the picture happen very suddenly or abruptly?
Time	Did the situation concern an event or an action that had happened in the past, that had just happened or that was to be expected for the future?
Pleasantness	This type of event, independent of your personal evaluation, would it be generally considered as pleasant or unpleasant?
Relevance	Was the event relevant for your general well-being, for urgent needs you felt, or for specific goals or plans you were pursuing at the time?
Expectedness	Did you expect the event and its consequences before the situation actually happened?
Conduciveness	Did the event help or hinder you in satisfying your needs, in pursuing your plans or in attaining your goals?
Urgency	Did you feel that action on your part was urgently required to cope with the event and its consequences?
Ego Cause	Was the event caused by your own actions - in other words, were you partially or fully responsible for what happened?
Other Cause	Was the event caused by one or several other persons - in other words, were other people partially or fully responsible for what happened?
Chance Cause	Was the event mainly due to chance?
Control	Can the occurrence and the consequences of this type of event generally be controlled or modified by human action?
Power	Did you feel that you had enough power to cope with the event - i.e. being able to influence what was happening or to modify the consequences?
Adjustment	Did you feel that, after having used all your means of intervention, you could live with the situation and adapt to the consequences?
Ext. Norm	Would the large majority of people consider what happened to be quite in accordance with social norms and morally acceptable?
Int. Norm	If you were personally responsible for what happened, did your action correspond to your self-image?

Each participant was invited to take place in front of a personal computer located in a lab room, and received an introduction, by the experimenter, on what was

to be expected and were asked to provide their informed-consent. The experimenter remained available for questions, but retreated to another part of the lab room such that the participant could fill in the questionnaire privately. In this experiment, each participant answered the 18 questions for five different, randomly chosen photos. The experiment took 25-30 minutes to complete. At the end of the experiment participants were debriefed.

6.4.2 Web experiment

Participants

The web-based experiment resulted in the acquisition of 79 completed online surveys (out of 436 invitations). Another 129 respondents completed the survey only partially. Often the incomplete sets of answers contained complete subsets corresponding to the appraisals of individual pictures. We compared these subsets with the appraisals from completed surveys and found that they were not statistically different and thus sufficiently similar to be included in the data set. Therefore we selected all (picture-wise) complete appraisals with the restriction that not all answers were indicated as "not applicable". This resulted in 261 appraisals, originating from 102 participants (74 male; overall age range 25-64 years, mean age 37) who appraised on average 2.6 pictures.

Stimuli

The set of 28 photos of the lab experiment were also used for the web experiment. Again, for each subject a number of photos were chosen randomly (with the restriction that no photo should be chosen more than once per subject) in such a way that over all participants all photos were presented an approximately equal number of times. In order to reduce the time required for completion of the questionnaire, the number of photos per participant was limited to three.

Procedure

With the exception of the number of photos shown (three in the web-based experiment versus five in the lab experiment), the procedure from the lab experiment was copied as much as possible to the web-based experiment, where the oral instructions of the lab experiment were presented as written text on the screen. In total the survey took 15-20 minutes to complete. LimeSurvey, as a back-end of the `www.simplicitylabs.net` environment, was used as environment to perform the survey in.

6.4.3 Classification Techniques

Primary focus of our analysis is the classification task of providing the correct emotion label given an appraisal vector. For this task we selected seven techniques of different nature: k-Nearest Neighbors, Artificial Neural Network, Support Vector Machine (linear and Radial Basis Function (RBF) kernel, Generalized Learning Vector Quantization, Generalized Matrix Learning Vector Quantization, Robust Soft Learning Vector Quantization, and class conditional means. The techniques have been described in detail in Chapter 1. The method of using class conditional means as prototypes will be used as reference technique as it was applied by Scherer (1993).

6.4.4 Data analysis

First we assessed whether the results of both experiments (lab and web) were sufficiently similar to be grouped into a single data set by performing t-tests per emotion per appraisal dimension. Furthermore, several of the below outlined analyses were performed on both datasets in separation.

The analyses proceeded as follows. In order to verify proper use of the scales used to measure the appraisal dimensions, we inspected the distributions of answers given per appraisal question, and compared our results with those found in previous work. After that, we analyzed the structure of the data and the discriminative strength of appraisal in terms of different emotions by applying Principal Component Analysis (PCA). Finally, we applied the data in a classification task and compared performance of several classification methods.

Comparison to previous work

With small exceptions in setup, our study was quite similar to the work of Scherer (1993). We determined whether this is also reflected in the results by applying t-tests on the means and standard deviations from our experiment and those in Scherer (1993) that were based on approximately 200 participants scoring the 15 appraisal dimensions based on recollecting a personal emotional memory. For each combination of emotion and appraisal dimension we performed a Student's t-test from the means and standard deviations. In a recent publication (Meuleman and Scherer 2013) performed similar analysis of classifying appraisal vectors into emotions, again from recollection of emotional events. They, however, only used 12 emotions classes and expanded the set of appraisal dimensions to 25 and explicitly and extensively included pairwise and three-way interactions.

PCA analysis and reduction of dimensions

We carried out a PCA (including several rotational variants) to investigate whether we needed all fifteen appraisal dimensions to explain most of the variance in the data from the questionnaires. The principal components found, we compared to the dimensions Valence, Arousal, and Dominance of Bradley and Lang (1994). The dimensions of this dimensional emotion model were inspired by the dimensional Pleasure, Arousal, Dominance (PAD) model by Russell and Mehrabian (1977), which has originally been found through component analysis.

Classification

Primary focus of our analysis is the classification task of providing the correct emotion label as indicated by the participant, given an appraisal vector. For this task we selected classification techniques of different nature (see e.g., Duda et al. (2000)): kNN, ANN, LVQ variants (Witoelar et al. 2011), and class conditional means. The method of using class conditional means as prototypes will be used as reference technique as it was applied by Scherer (1993). For all distance based methods we used squared Euclidean distance as distance measure. LVQ classifiers were chosen because they are of open box nature, that is, they provide direct insight into the information learned by the classifier, as prototypes that are represented in the same space as the data can be inspected (similar to class conditional means).

As measures of performance, we report both Accuracy and Cohen's Kappa, the two most suitable measures for multi-class classification. Note that Cohen's Kappa was selected because it explicitly takes into account imbalance in the dataset.

All classifiers were applied using 10-fold participant-wise cross validation (i.e., all data from a subject is either used for training or for validation) used and the results reported are averages over 10×10-fold cross validations. As a baseline reference we used a baseline classifier that returned the label of the class with highest prior probability, as specified by Equation (6.1). It describes the best guess one can do without inferring any knowledge about the structure of the data, nevertheless it corrects for unbalanced class sizes.

$$\gamma = \arg\max_c p_c, \text{ where } p_c = \frac{n_c}{\sum_{\tilde{c}} n_{\tilde{c}}}, \tag{6.1}$$

where p_c is the class prior, inferred from the number of samples n_c representing class c.

In order to address imbalance in the dataset, we also trained classifiers on subsets of data labeled with emotions that had been rated at least 20 times, 10 times, and 5

times, leaving out underrepresented classes. Finally we trained classifiers using only the 'basic' emotions (Anger, Disgust, Fear, Joy, and Sadness) of Ekman (1972). For each of these classification tasks we used all 15 appraisal dimensions as input; the number of data samples used and class labels predicted will be listed in the table with results.

We optimize for testing accuracy (percentage of correctly classified samples), and next to that, will report the accuracy taking not only the best hit, but also second hit into account (such that direct comparison with the work of Scherer (1993) is possible). Note that for SVM we did not include second hit performance because the SVM implementation we used does not offer that functionality and could not easily be extended.
We analyzed the confusion matrices to study whether particular emotions being classified incorrectly were responsible for the classifiers' performance, or whether classification errors were more generally spread over the classes.

Finally, we used the inherent transparency of prototype-based learning by assessing the (relative) positions of the prototypes to get a better understanding of how the data was structured.

Table 6.2: *Properties of collected data, means and standard deviations per emotion, per appraisal dimension.*

	Novelty	Time	Pleasantness	Relevance	Expectedness	Conducive-ness	Urgency	Ego Cause
Enjoyment	1.86 ± 0.95	2.50 ± 0.93	4.50 ± 0.76	1.58 ± 1.01	3.39 ± 1.59	3.14 ± 1.29	1.67 ± 1.15	2.58 ± 1.79
Joy	1.93 ± 1.29	2.36 ± 0.98	4.32 ± 1.04	1.66 ± 1.35	3.18 ± 1.63	2.93 ± 1.42	1.59 ± 1.40	2.02 ± 1.67
Disgust	2.16 ± 1.52	2.07 ± 1.04	1.28 ± 0.54	1.40 ± 1.20	2.33 ± 1.62	1.79 ± 1.36	1.86 ± 1.50	1.21 ± 0.73
Contempt	2.08 ± 1.14	2.62 ± 0.49	1.62 ± 0.84	1.23 ± 0.70	2.15 ± 1.56	2.31 ± 1.38	1.38 ± 1.15	2.00 ± 1.52
Sadness	2.63 ± 1.66	2.17 ± 1.05	1.37 ± 0.72	2.06 ± 1.51	2.38 ± 1.55	1.89 ± 1.31	1.89 ± 1.26	1.22 ± 1.05
Desperation	2.90 ± 1.62	2.17 ± 0.97	1.33 ± 0.65	2.53 ± 1.89	2.07 ± 1.12	1.37 ± 1.11	2.53 ± 1.80	1.83 ± 1.19
Anxiety	2.42 ± 1.43	2.10 ± 1.20	1.84 ± 1.19	2.26 ± 1.48	2.97 ± 1.64	2.42 ± 1.52	2.03 ± 1.40	1.87 ± 1.50
Fear	3.00 ± 1.67	2.26 ± 1.18	1.64 ± 1.18	2.57 ± 1.71	2.77 ± 1.47	1.68 ± 1.50	2.30 ± 1.70	2.02 ± 1.49
Irritation	2.48 ± 1.47	2.24 ± 0.92	2.10 ± 1.11	1.38 ± 1.00	2.81 ± 1.40	2.29 ± 1.12	2.00 ± 1.57	1.95 ± 1.43
Anger	2.63 ± 1.49	2.50 ± 0.87	1.75 ± 0.97	2.38 ± 1.58	2.83 ± 1.55	1.75 ± 1.39	2.04 ± 1.57	1.17 ± 1.14
Indifference	2.22 ± 1.34	2.34 ± 1.21	2.28 ± 0.84	1.88 ± 1.41	2.81 ± 1.76	2.59 ± 1.45	1.53 ± 1.12	1.31 ± 1.10
Shame	2.54 ± 1.15	2.38 ± 1.27	1.85 ± 0.77	2.15 ± 1.51	2.38 ± 1.08	1.69 ± 1.20	1.85 ± 1.70	1.46 ± 1.34
Guilt	3.50 ± 1.38	2.33 ± 0.47	2.00 ± 0.58	2.00 ± 0.82	3.67 ± 1.25	2.50 ± 0.76	2.33 ± 1.37	2.33 ± 1.11
Pride	1.93 ± 1.03	2.21 ± 0.94	4.29 ± 1.22	2.64 ± 1.39	3.07 ± 1.87	3.57 ± 1.35	2.36 ± 1.63	2.86 ± 1.77

	Other Cause	Chance Cause	Control	Power	Adjustment	Ext. Norm	Int. Norm
Enjoyment	3.11 ± 1.59	1.83 ± 1.04	3.61 ± 1.30	2.86 ± 1.67	3.78 ± 1.80	4.08 ± 1.09	3.00 ± 1.83
Joy	2.57 ± 1.74	2.16 ± 1.33	3.45 ± 1.23	2.36 ± 1.73	3.59 ± 2.08	4.11 ± 1.48	2.57 ± 1.94
Disgust	4.28 ± 1.34	1.63 ± 1.10	4.02 ± 1.21	1.65 ± 1.12	2.74 ± 1.63	1.60 ± 0.97	1.00 ± 0.81
Contempt	4.23 ± 0.89	1.23 ± 0.58	4.46 ± 0.84	2.38 ± 1.39	2.85 ± 1.51	1.69 ± 1.14	1.54 ± 1.01
Sadness	2.75 ± 1.82	2.11 ± 1.31	3.11 ± 1.44	1.94 ± 1.23	2.75 ± 1.46	2.48 ± 1.59	1.54 ± 1.40
Desperation	2.77 ± 1.61	2.27 ± 1.46	3.30 ± 1.27	1.33 ± 1.07	2.20 ± 1.38	1.90 ± 1.72	1.73 ± 1.46
Anxiety	2.71 ± 1.87	1.74 ± 1.24	3.74 ± 1.24	2.29 ± 1.44	3.13 ± 1.98	3.10 ± 1.55	2.42 ± 1.91
Fear	2.17 ± 1.65	2.51 ± 1.46	2.98 ± 1.43	2.08 ± 1.15	2.60 ± 1.47	1.77 ± 1.69	1.49 ± 1.59
Irritation	3.95 ± 1.40	1.95 ± 1.00	3.71 ± 1.20	2.48 ± 1.37	3.33 ± 1.43	2.38 ± 1.29	1.76 ± 1.15
Anger	3.88 ± 1.64	1.25 ± 0.78	3.79 ± 1.38	1.88 ± 1.45	2.13 ± 1.36	1.96 ± 1.37	1.79 ± 1.35
Indifference	3.19 ± 1.74	2.00 ± 1.35	3.56 ± 1.25	1.81 ± 1.36	3.56 ± 1.78	2.94 ± 1.66	1.34 ± 1.55
Shame	2.54 ± 1.74	2.15 ± 0.86	3.92 ± 1.07	1.92 ± 1.54	3.08 ± 1.82	2.62 ± 1.55	1.77 ± 1.05
Guilt	2.50 ± 0.76	2.67 ± 0.75	4.00 ± 0.58	2.50 ± 1.26	4.00 ± 0.58	3.50 ± 0.50	2.50 ± 1.12
Pride	4.36 ± 1.04	1.57 ± 0.73	4.43 ± 0.62	2.50 ± 1.80	3.50 ± 2.03	4.00 ± 1.41	3.79 ± 1.61

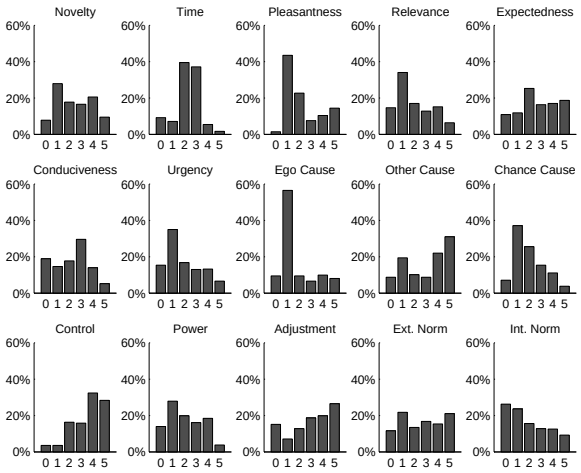

Figure 6.3: *The distributions of answers given per appraisal dimension.*

6.5 Results

First we compared the data originating from the lab and web experiment. t-tests per emotion per appraisal dimension indicated only 4 (out of $14 * 15 = 210$) significant differences ($p < 0.05$) between the two experiments. We also compared the outcomes of PCA analysis on both datasets in isolation, which showed that the two data sets were highly similar. In fact similar to such an extent that they could be merged, resulting in a data set containing 420 appraisals obtained from 134 individual participants, summarized in Table 6.2. The t-tests to compare our dataset with that in Scherer (1993) revealed no significant ($p < 0.05$) differences, only one marginally insignificant difference ($p = 0.055$) was found for (Joy, Ext. Norm.).

We inspected the answers given per appraisal question in order to verify that the scales are used as expected. The distributions per appraisal question are depicted in Figure 6.3, normalized for the number of answers. Most distributions are fairly uniformly distributed. The distribution of Control is skewed to the right, Pleasantness to the left, as is Ego Cause, and Time shows a larger contribution of the mid-section of the range. We observe that the scales were completely used rather than just the outer ends of the scale. In the following analyses we will interpret the data as interval data.

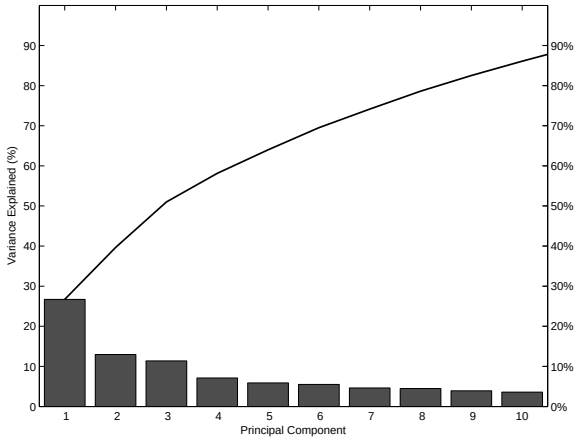

Figure 6.4: *The percentage of variance explained per principal component (bars) and total percentage of variance with each additional component (line).*

6.5.1 Component analysis

The PCA revealed that the first principal component (i.e., the direction in which the variance of the data is largest) described 27% of the data and that the first three components sum up to 51%. In order to get above 80% of variance explained, the first 9 principal components were needed (see Figure 6.4).

Projecting the original (appraisal) dimensions onto the dimensions found by PCA, as depicted in Figure 6.5, provided insight in the interrelations of appraisal dimensions. The length and orientation of the line segments indicate how much the appraisal dimension contributed to the first and second PCA components. For instance, appraisal dimensions Conduciveness and Expectedness were closely aligned along the first component. Other combinations of appraisal dimensions that have similar contributions to the first two principal components are Ego Cause and Power, Adjustment and Internal Normative Significance, and Relevance and Urgency. It also shows that Novelty and Causality by Chance were rather opposite to Control. Time is less prominently represented in the first two principal components.

Because components 1-3 collectively explained a little over half of the variance, we tried to label these components to come to a better understanding of how the data was structured. An initial visual inspection hinted that components 1-3 could perhaps be related to the Valence, Arousal, and Dominance dimensions of Bradley and Lang (1994). In an attempt to label the principal components found in our dataset, we tested this hypothesis.

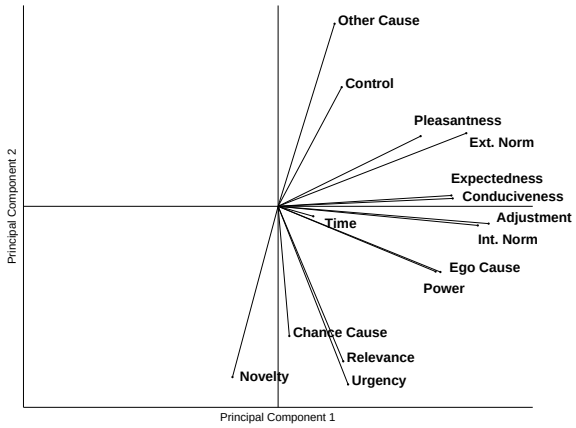

Figure 6.5: *Projection of the appraisal dimensions into the space spanned by principal components 1 and 2.*

Table 6.3: *Correlations between the first three principal components and affective dimensions (*$p < 0.01$, **$p < 0.001$).*

Principal components \ Affective dimensions	Valence	Arousal	Dominance
1	** 0.92	0.39	* 0.60
2	-0.45	-0.07	** -0.71
3	-0.16	-0.05	* -0.55

We asked a new group of participants ($N = 29$) to rate each of the 14 emotion words on a Valence, Arousal, Dominance (VAD) scale and used these data to assess whether any of the three components would correlate to one of these affective dimensions. These participants did not participate in the experiment described in section 6.4, but had similar education level, age, and gender distribution. The instructions used can be found in Table 6.4. We calculated correlations between the three components and the three affective dimensions (see Table 6.3) and observed that the first principle component correlates very highly ($\rho = 0.92$) with valence and

the second component correlates with dominance ($\rho = -0.71$).[1] Dominance was also present (though with less strength) in the first and third principal component. Arousal was not found to correlate with any of the first three principal components.

Valence is known to be related to conduciveness and intrinsic pleasure (Broekens 2012). From Figure 6.5 we observe that these appraisal dimensions are present in the first principal component which highly correlates with valence. Similarly, Dominance is known to be related to be related to Power-to-cope and Control (Broekens 2012), which are present in both first and second principal component that both correlate with dominance.

Table 6.4: *The instructions used to assess Valence, Arousal and Dominance.*

Please rate the following emotion words on valence, arousal and dominance, using numbers [1..10]. It is ok to give multiple emotions the same number. These scales are defined as follows:

- Low valence means that the emotion feels negative;
- High valence means that the emotion feels positive.
- Low arousal means that the emotion does not excite you;
- High arousal means that the emotion does excite you very much.
- Low dominance means that you do not feel in control of the emotion;
- High dominance means that you do feel in control of the emotion.

[1] Note that due to an administrative error different results were obtained in earlier analysis performed and reported about in de Vries et al. (2009).

Table 6.5: *Test performance averaged over* 10×10*-fold participant-wise cross validations for the complete set of 14 classes and specified subset of 10 classes.*

	Complete dataset: 14 classes				Rated 20+ times: 10 classes			
Method	Hyper-parameter	Accuracy	Accuracy incl. 2nd hit	Kappa	Hyper-parameter	Accuracy	Accuracy incl. 2nd hit	Kappa
Baseline		14.9%	27.4%	0.06		16.7%	30.8%	0.06
kNN	$k = 9$	$23.1\% \pm 1.6\%$	$39.2\% \pm 2.1\%$	0.14 ± 0.02	$k = 7$	$26.6\% \pm 2.3\%$	$43.3\% \pm 2.3\%$	0.17 ± 0.03
ANN	$N_{hidden} = 3$	$23.4\% \pm 2.1\%$	$41.4\% \pm 1.9\%$	0.13 ± 0.02	$N_{hidden} = 3$	$25.8\% \pm 2.6\%$	$46.0\% \pm 2.8\%$	0.14 ± 0.03
SVM - linear	$C = 0.1$	$25.6\% \pm 1.7\%$	$47.0\% \pm 1.4\%$	0.17 ± 0.02	$C = 0.01$	$28.7\% \pm 2.1\%$	$52.0\% \pm 2.2\%$	0.18 ± 0.02
Means		$21.1\% \pm 1.5\%$	$40.2\% \pm 2.5\%$	0.15 ± 0.02		$26.6\% \pm 1.7\%$	$48.5\% \pm 1.9\%$	0.18 ± 0.02
GLVQ		$22.3\% \pm 1.9\%$	$38.7\% \pm 1.3\%$	0.14 ± 0.02		$25.3\% \pm 1.9\%$	$43.7\% \pm 3.1\%$	0.16 ± 0.02
RSLVQ	$v_{soft} = 0.5$	$24.7\% \pm 0.8\%$	$45.7\% \pm 2.1\%$	0.18 ± 0.01	$v_{soft} = 10$	$28.0\% \pm 2.1\%$	$51.1\% \pm 1.7\%$	0.20 ± 0.02
GMLVQ		$14.8\% \pm 1.5\%$	$27.4\% \pm 2.4\%$	0.06 ± 0.01		$18.1\% \pm 2.0\%$	$34.7\% \pm 2.0\%$	0.07 ± 0.02

Table 6.6: *Test performance averaged over* 10×10*-fold participant-wise cross validations for specified subsets of 5 classes.*

	Rated 35+ times: 5 classes				Basic Emotions: 5 classes			
Method	Hyper-parameter	Accuracy	Accuracy incl. 2nd hit	Kappa	Hyper-parameter	Accuracy	Accuracy incl. 2nd hit	Kappa
Baseline		26.4%	48.5%	0.07		27.8%	51.1%	0.08
kNN	$k = 7$	$48.1\% \pm 2.5\%$	$77.4\% \pm 1.9\%$	0.34 ± 0.03	$k = 11$	$51.4\% \pm 2.6\%$	$69.1\% \pm 1.9\%$	0.37 ± 0.03
ANN	$N_{hidden} = 3$	$47.0\% \pm 3.1\%$	$80.2\% \pm 2.8\%$	0.32 ± 0.04	$N_{hidden} = 3$	$49.3\% \pm 2.8\%$	$69.9\% \pm 2.5\%$	0.34 ± 0.03
SVM - linear	$C = 0.01$	$51.8\% \pm 4.0\%$	$86.4\% \pm 1.8\%$	0.39 ± 0.05	$C = 1$	$54.5\% \pm 3.6\%$	$76.3\% \pm 2.1\%$	0.42 ± 0.05
Means		$49.7\% \pm 4.0\%$	$81.6\% \pm 2.8\%$	0.37 ± 0.05		$49.7\% \pm 3.0\%$	$72.9\% \pm 3.3\%$	0.37 ± 0.04
GLVQ		$48.1\% \pm 3.3\%$	$76.4\% \pm 3.1\%$	0.35 ± 0.04		$51.6\% \pm 3.2\%$	$73.7\% \pm 3.6\%$	0.38 ± 0.04
RSLVQ	$v_{soft} = 10$	$48.7\% \pm 3.6\%$	$82.8\% \pm 2.1\%$	0.36 ± 0.04	$v_{soft} = 1$	$54.3\% \pm 3.1\%$	$77.1\% \pm 2.6\%$	0.42 ± 0.04
GMLVQ		$46.2\% \pm 3.0\%$	$77.7\% \pm 2.9\%$	0.32 ± 0.04		$46.3\% \pm 3.1\%$	$67.7\% \pm 4.1\%$	0.31 ± 0.04

Table 6.7: *Training performance averaged over* 10×10*-fold participant-wise cross validations for the complete set of 14 classes and specified subset of 10 classes.*

	Complete dataset: 14 classes				Rated 20+ times: 10 classes			
Method	Hyper-parameter	Accuracy	Accuracy incl. 2nd hit	Kappa	Hyper-parameter	Accuracy	Accuracy incl. 2nd hit	Kappa
Baseline		14.9%	27.4%	0.06		16.7%	30.8%	0.06
kNN	$k = 9$	$49.9\% \pm 0.5\%$	$63.8\% \pm 0.4\%$	0.44 ± 0.01	$k = 7$	$55.8\% \pm 0.4\%$	$69.4\% \pm 0.5\%$	0.50 ± 0.00
ANN	$N_{hidden} = 3$	$34.7\% \pm 0.6\%$	$54.1\% \pm 1.0\%$	0.25 ± 0.01	$N_{hidden} = 3$	$39.8\% \pm 0.6\%$	$61.6\% \pm 1.1\%$	0.30 ± 0.01
SVM - linear	$C = 0.1$	$44.8\% \pm 0.5\%$	$60.0\% \pm 0.7\%$	0.38 ± 0.01	$C = 0.01$	$37.3\% \pm 0.4\%$	$59.3\% \pm 0.4\%$	0.27 ± 0.01
Means		$31.6\% \pm 0.6\%$	$51.4\% \pm 0.4\%$	0.26 ± 0.01		$36.0\% \pm 0.6\%$	$57.5\% \pm 0.4\%$	0.29 ± 0.01
GLVQ		$40.2\% \pm 1.5\%$	$53.0\% \pm 1.0\%$	0.34 ± 0.02		$36.7\% \pm 0.8\%$	$55.6\% \pm 0.6\%$	0.28 ± 0.01
RSLVQ	$v_{soft} = 0.5$	$40.7\% \pm 0.7\%$	$58.6\% \pm 0.7\%$	0.35 ± 0.01	$v_{soft} = 10$	$39.5\% \pm 0.6\%$	$60.5\% \pm 0.3\%$	0.32 ± 0.01
GMLVQ		$17.9\% \pm 1.1\%$	$30.7\% \pm 1.3\%$	0.09 ± 0.01		$22.8\% \pm 1.1\%$	$37.8\% \pm 1.4\%$	0.13 ± 0.01

Table 6.8: *Training performance averaged over* 10×10*-fold participant-wise cross validations for specified subsets of 5 classes.*

	Rated 35+ times: 5 classes				Basic Emotions: 5 classes			
Method	Hyper-parameter	Accuracy	Accuracy incl. 2nd hit	Kappa	Hyper-parameter	Accuracy	Accuracy incl. 2nd hit	Kappa
Baseline		26.4%	48.5%	0.07		27.8%	51.1%	0.08
kNN	$k = 7$	$66.0\% \pm 0.7\%$	$88.9\% \pm 0.4\%$	0.57 ± 0.01	$k = 11$	$66.0\% \pm 0.3\%$	$82.3\% \pm 0.3\%$	0.56 ± 0.00
ANN	$N_{hidden} = 3$	$66.3\% \pm 1.2\%$	$92.4\% \pm 0.7\%$	0.57 ± 0.02	$N_{hidden} = 3$	$70.2\% \pm 1.5\%$	$85.7\% \pm 0.8\%$	0.61 ± 0.02
SVM - linear	$C = 0.01$	$57.8\% \pm 0.4\%$	$88.2\% \pm 0.2\%$	0.46 ± 0.01	$C = 1$	$73.6\% \pm 0.5\%$	$87.4\% \pm 0.8\%$	0.66 ± 0.01
Means		$54.1\% \pm 0.5\%$	$83.0\% \pm 0.3\%$	0.43 ± 0.01		$55.7\% \pm 0.5\%$	$76.6\% \pm 0.4\%$	0.44 ± 0.01
GLVQ		$58.1\% \pm 0.5\%$	$80.8\% \pm 0.4\%$	0.48 ± 0.01		$60.4\% \pm 0.6\%$	$78.3\% \pm 0.7\%$	0.50 ± 0.01
RSLVQ	$v_{soft} = 10$	$57.5\% \pm 0.5\%$	$85.2\% \pm 0.4\%$	0.47 ± 0.01	$v_{soft} = 1$	$65.6\% \pm 0.5\%$	$84.7\% \pm 0.5\%$	0.57 ± 0.01
GMLVQ		$60.9\% \pm 1.5\%$	$81.0\% \pm 1.4\%$	0.51 ± 0.02		$60.0\% \pm 2.0\%$	$75.5\% \pm 1.8\%$	0.48 ± 0.02

6.5.2 Classification

We trained various classifiers, and collected their performances in Table 6.5, which shows that the accuracy of the classifiers ranged from 14.4% to 25.6% on the complete data set. Corresponding Cohen's Kappa values range from 0.06 to 0.18. The baseline classifier obtained 14.9% accuracy. The linear Support Vector Machine (RBF kernel did not improve upon the linear kernel) outperforms the other classifiers in terms of accuracy, followed by RSLVQ and ANN. Using class-conditional means as prototypes only reaches 21.1% accuracy. GMLVQ seems to struggle with this classification task. When also considering second-hit a success, we observe performances between 25.5% and 47.0%, where the baseline classifier reached 27.4% accuracy. Based upon Cohen's Kappa, we observe that RSLVQ reaches best performance ($\kappa = 0.18$), followed by the SVM and class conditional means.

The data set showed non-uniform priors, indicating that some classes were underrepresented. Therefore, the classifiers might not be able to learn the properties of these small classes to a sufficient degree. To reduce this effect, we focused on the data samples labeled with emotions that had been rated at least 20 times, containing 10 emotion classes, as well as the subset of emotions that had been rated at least 35 times, consisting of 5 emotion classes (Enjoyment, Joy, Disgust, Sadness, and Fear). For the 10-class classification, results are reported in the sixth column of Table 6.5. All classifiers are able to reach higher performances on this classification task, and RSLVQ almost catches up with the SVM, which is still leading (with 28.0% and 28.7%, respectively). In terms of Cohen's Kappa, RSLVQ is still leading, now with $\kappa = 0.20$. Reducing the data set further to only the top 5 most rated classes (rated at least 35 times) showed a large increase of accuracy (third column of Table 6.6), up to 51.8% ($\kappa = 0.39$). The baseline performance also increased significantly (to 26.4%) because of the reduced number of classes. GMLVQ has caught up with the other methods and shows only shows slightly worse performance than the other LVQ methods. When classifying only 'basic' emotions (Anger, Disgust, Fear, Joy, and Sadness) of Ekman (1972), we found similar results, where SVM and RSLVQ reached top accuracies of 54.5% and 54.3%, respectively, and $\kappa = 0.42$.

In all settings there were at least 227 data samples used for training/validation, and RSLVQ proved to be the best classifier for this task, usually followed by SVM and ANN. Tables 6.7 and 6.8 show the training performances of the classifiers on the four different classification tasks. The relatively small differences between test and training performances show that none of the classifiers is prone to severe overfitting.

When investigating the confusion matrices, we found the ANN to discard 4 of

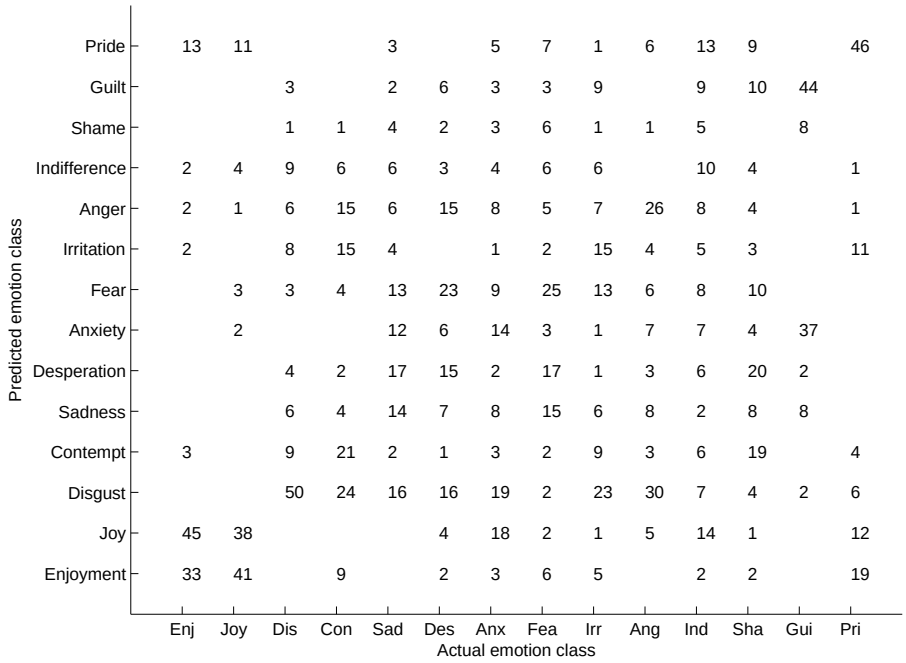

Figure 6.6: *Confusion matrix showing the (participant indicated) labeled emotions on the horizontal axis and the (RSLVQ) classified instances vertically, when considering all emotions. The numbers represent the percentages of classified emotions per labeled emotion.*

the 15 classes (it never assigned these labels to samples in the test set). Figure 6.6 shows the confusion matrix of RSLVQ (applied to the data of all 14 emotions). It revealed that for most emotions at least some instances are classified correctly. However, no single instance of Shame was classified as such, indicating that Shame was easily mistaken for other emotions. Joy and Enjoyment were confused often, as well as Disgust and Contempt. Fear turned out to be confused often with Desperation and Sadness. Confusion matrices of the other classifier (apart from ANN) showed high resemblance of the one depicted.

The classification task using only 'basic' emotions separated out these easily confused emotions, which resulted in higher classification performance, and a confusion matrix as shown in Figure 6.7. The confusion matrix clearly shows best recognition of the emotion Joy, while Anger, Sadness, and Fear are more often confused.

We took the learned prototypes of RSLVQ and projected them in earlier found principal component space. Figure 6.8 shows that the prototypes settled at mostly

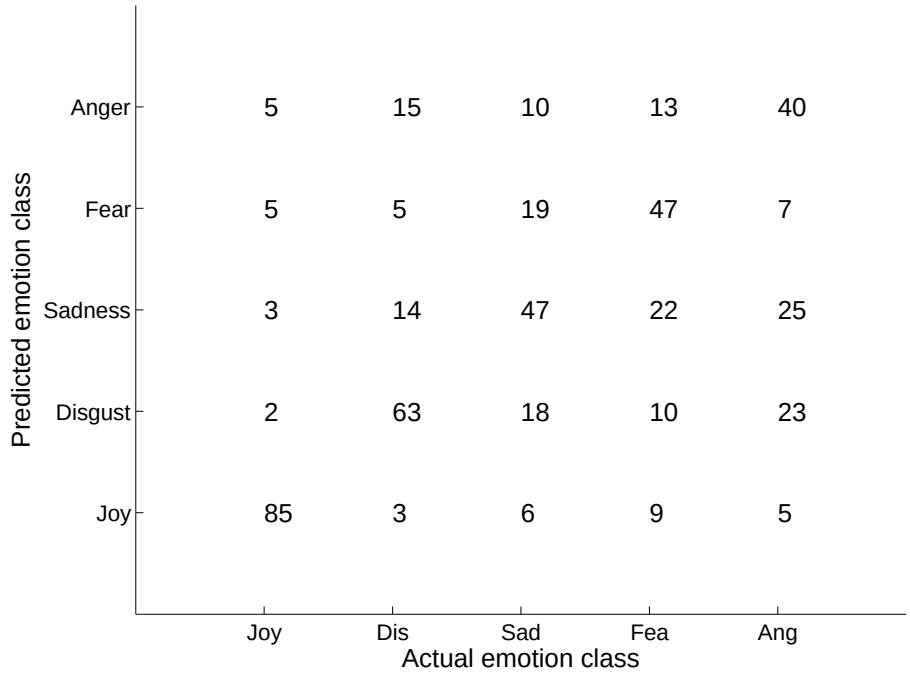

Figure 6.7: *Confusion matrix showing the (user indicated) labeled emotions on the horizontal axis and the (RSLVQ) classified instances vertically, when considering only 'basic' emotions. The numbers represent the percentages of classified emotions per labeled emotion.*

non-coinciding locations and showed a distinct grouping of positive emotions (Enjoyment, Joy, and Pride) separate from the negative emotions. Repeated tries (using different subsets of data for training the prototypes) resulted in similar graphs. We did find some small variations in position amongst the negative emotions.

6.6 Discussion

We proposed an alternative method to the gold-standard of a one-step emotion classification by considering the process of mapping measurements to emotion labels as a two-stage approach in which first a mapping to an appraisal representation is made, followed by a second step in which the appraisal is mapped onto emotion labels. We gathered data to determine to what extent an appraisal model (here the CPM) can be used in such an approach by investigating the second step in this two-stage process. We collected data from 134 participants who were asked to empathize

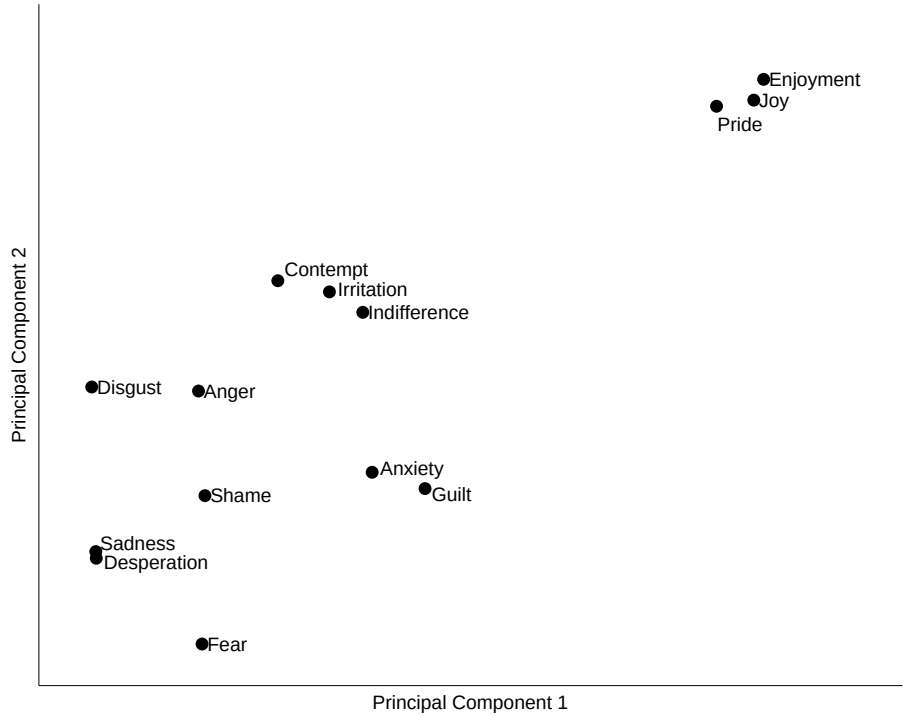

Figure 6.8: *RSLVQ prototypes projected in a reduced (2D) space, using the earlier found PCA projection.*

with photos shown and rate the emotions they experienced as well as providing ratings to the questions relating to Scherer's 15 appraisal dimensions (Scherer 1993). No significant differences were found in the t-tests performed to compare our dataset to that in previous work (Scherer 1993), indicating that we successfully replicated the original experiment (using recollection to elicit emotions) in a different setting, using empathic photos.

We found that over 90% of the variance in our data could be explained by the first 9 principal components. The Principal Component Analysis, confirmed expected interdependencies between the appraisal dimensions. Within the first two components (explaining over 40% of variance), for instance, appraisal dimensions Conduciveness and Expectedness were closely aligned along the first component. This can be explained since unexpected events are usually not conducive regarding one's goals. Similar influence of Ego Cause and Power can be explained by the

observation that one likely feels power over self-caused actions and resulting consequences. Adjustment and Internal Normative Significance are related through the observation that one can live with and adapt to consequences originating from actions that are in line with one's internal norms. Relevance and Urgency are linked through the observation that events that require urgent action must be relevant. The PCA also shows that Novelty and Causality by Chance were rather opposite to Control. Also this can be explained by people not feeling (fully) in control when new and/or seemingly random events happen. These observations do not take into account all subtleties taken into account in the design of the CPM model; nevertheless the results of the PCA might indicate that, given the circumstances of our experiment, the CPM appraisal model could be simplified by combining certain pairs of appraisal dimensions.

The PCA and correlation analysis also indicate that the first three principal components map well onto the affective dimensions valence and dominance, which is in line with previous work (Scherer et al. 2006, Broekens 2012) in which VAD scores were compared to a manually chosen linear combination of appraisal dimensions. The strongest connection that we observe is one between the first principal component and valence with similar strength as found in the former study. We observe that arousal is less present in the principal components, which might follow from less emotional strength felt through empathizing compared to other ways of emotion elicitation like recall techniques.

When applying the data in a classification task, results show that, when using all 14 emotion classes, classifying performance reached close to 25% accuracy and a kappa of 0.18. Following the guideline of Landis and Koch (1977), this should be interpreted as a 'slight agreement' between the classifiers output and ground truth. One should, however take into account that our tasks involves 14 classes, which is substantially more than the 4 classes the guideline was devised for. The RSLVQ classifier produced the highest classification performance for all subtasks including the 14-class classification and the chosen (and more balanced) subtasks involving 10 and 5 emotion classes. Highest performance (54.3% accuracy; kappa: 0.42) was observed in the task of classifying into the 5 'basic' emotions of Ekman (1972). Using the aforementioned guideline for the interpretation of the kappa value, we can consider this as 'moderate agreement'.

As discussed in Section 6.3, there is limited work to compare these results with. Scherer also worked on a classifier of appraisal vectors onto emotion labels in the 'Geneva Expert System on Emotion' (GENESE) (Scherer 1993), and reports an accu-

racy of 78%. There are, however, three aspects that need to be taken into account when comparing the results. First, Scherer's system considered both the first- and second-best guess as correct. Second, almost 10% of data was excluded in GENESE (based on the criterion of differing more than 0.5 standard deviations). Third, users were asked whether the prediction "possibly reflects some part of what they felt in the situation, possibly without realizing it", which artificially increases dependency between appraisal data and emotion labels due to the bias of participants to confirm experiment hypotheses, referred to as the good-subject effect (see e.g., Nichols and Maner (2008)). Our results are on par with the performance reported on 12-emotion classification by Meuleman and Scherer (2013) of 27.9%, which used 10 more appraisal dimensions. Unfortunately they don't report details on the unbalance present in their dataset such that we cannot further compare the results.

In order to benchmark our system as part of the proposed two-stage emotion classification, we decide to compare against a human benchmark (Janssen, Tacken, de Vries, van den Broek, Westerink, Haselager and IJsselsteijn 2013) of classifying signals into 5 emotion classes. One should note that this benchmark considers the full process of labelling measurement data to emotions rather than just the second stage in our proposed two-stage approach. Unimodal 5-class emotion recognition by humans reached up to 26.1% accuracy which is approximately half the performance our system obtained (54.3%). Hence we outperform the benchmark with quite a margin. Considering that the performance of a full two-stage emotion classification system that uses our classifier in the second stage will have a performance of at most this 54.3%, we believe this margin in performance still leaves room to perform better than the human benchmark also on the integral task.

The differences we observed in how well certain emotions could be classified might originate from the elicitation method because some emotions could be harder to elicit through photo viewing than others. Another explanation could be that for some emotions more than others, there might still be interpersonal difference in the link between of appraisal and emotions. This would, however, be in contrast to the way appraisal models are proposed in literature.

The use of prototype based classifiers enables us to further explore the differences and similarities between emotions. It reveals that especially the positive and negative emotions can clearly be distinguished from one another, which is confirmed by analysis of the confusion matrices (Figure 6.6 and 6.7). A closer look at the projection of the prototypes in the principal component space (Figure 6.8) reveals some finer level grouping of emotions. We found Sadness and Desperation

very close together. This pair of emotions is known as a low/high intensity pair (Baldaro et al. 2003), or in other theories as derived emotions from to the primary emotion Sadness (Shaver et al. 2001). The closeness and ordering of Indifference, Irritation and Contempt might suggest a trio of similar emotions in terms of intensity. Irritation and Contempt are seen as secondary and tertiary emotions of Anger, and Douglas-Cowie et al. (2006) list Irritation and Contempt both in the category of ″Negative & forceful″ emotions. This might explain the closeness of Irritation and Contempt. Our findings suggest that Indifference might be related to these two emotions as a weaker form of the same base emotion. Comparing our work with the clusters found in Meuleman and Scherer (2013), our results confirm their finding of a clearly separated cluster of positive emotions as well as a cluster of anger emotions. Distress, shame and guilt emotions were clustered together. We did not find evidence for a further sub-cluster of shame/guilt within the latter cluster.

Our study has some limitations. First, we used photo-viewing as elicitation method. The limited presence of arousal in the first principal components indicates that the emotions experienced (and rated) by our participants were limited in strength. Also, despite the promising results from our classification analysis, the observed performances leave room for improvement, indicating that the classes showed relatively large amounts of overlap in the data (i.e., there was no clear separation between classes). These arguments indicate that due to relatively low strength of emotions felt, the appraisal data corresponding to different emotion classes might have been more similar than they could have been when other emotion elicitation methods would have been used. This observation is in line with the statement that empathic photo-viewing constitutes a tough paradigm for the use of appraisal criteria, as discussed in Scherer et al. (2006).

Second, the overlap in the data might originate (in part) from interpersonal differences among our participants. These interpersonal differences might originate from different interpretation of the questions and answering options of the CPM appraisal model. We have chosen to stay as close to the original questionnaire (Scherer 1993) as possible which presents uni- or bipolar answering options to yes-no questions. It might be that a better design of the questionnaire reduces the interpersonal variance. On the other hand, an alternative explanation of the interpersonal differences is that they could be inherent to the appraisal model. That would suggest that we might need to reconsider whether indeed the mapping of appraisals to emotion is a general process independent of the individual.

6.7 Conclusion

Our results indicate that the CPM generalizes from contexts of recollection of personal emotional events to empathizing with (generic) emotional events. Our attempt to discriminate between discrete emotions on the level of appraisal, in a pure classification task, exceeds baseline performance with a factor two. When focusing on the subset of the five 'basic' emotion classes from Ekman (1972), we obtain classification accuracy of up to 54.3% (Cohen's kappa: 0.42). Our classifier reaches approximately twice the performance of the human benchmark on a 5-class classification problem, leaving a large margin. Considering that the performance of a full two-stage emotion classification system that uses our classifier in the second stage will have a performance of at most this 54.3%, we believe this margin in performance still leaves room to perform better than the human benchmark also on the integral task.

In conclusion, we have shown a first attempt towards automated emotion classification using appraisal representations as an intermediate step in a two-stage process of emotion classification. In order to do so, we have evaluated the second step, classifying appraisals to emotions using machine learning techniques. Future work should address also the first step of the proposed two-step approach for affect classification, namely the automated extraction of appraisal vectors from affective measurements such as physiology, audio or video of a user. There have been established correlations for many appraisal dimensions and, e.g., physiological measures in a well-controlled environment and in isolation (Aue et al. 2007, Grandjean and Scherer 2008, Poli et al. 2007, Smith 1989); a full coverage of all appraisal dimensions in general circumstances still requires further attention.

Given that we used photo-viewing as an elicitation method, which is particularly tough for the application of appraisal modeling, we belief there is much more potential for the proposed system, and appraisal modeling in general, when tested with emotion eliciting paradigms that generate stronger emotions. It would be interesting, in a next step, to further evaluate this approach in such circumstances. While the two-step approach could, in principle, be applied in any emotion classification system, it might be most applicable to applications that have (to some extent) control over the stimuli that users are exposed to, such that stimuli-characteristics are most easily measured and can be used as contextual information to add to the measurements of the user. Applications in this area are dynamic adaptation to emotions by flight simulators, serious games, office suite software etc. In these circumstances emotion classification systems have the potential to adapt the way computers act and react to their users to create much more natural human-computer interaction, improve realism and reduce frustration.

Material based on:
de Waele, S., de Vries, G.-J. and Jager, M.: 2009, Experiences with adaptive statistical models for biosignals in daily life, Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on, pp. 1-6.
Westerink, J., van Beek, W., Daemen, E., Janssen, J., de Vries, G.-J. and Ouwerkerk, M.: 2014, The vitality bracelet: Bringing balance to your life with psychophysiological measurements, in S. H. Fairclough and K. Gilleade (eds), Advances in Physiological Computing, Springer London, London, pp. 197-209.

Hoofdstuk 7

APPLICATIONS

7.1 Introduction

Next to the research into classification of various affective states from a variety of measurement modalities, covered in Chapters 4, 5 and 6, this chapter explores various applications. Based upon the insights gathered we have developed demonstrator systems that operationalize results obtained. The aim in the remainder of this chapter is not to provide fully validated systems, but rather present several potential applications based upon the research performed.

7.2 Vitality Bracelet

In Chapter 4 we have explored three physiological modalities for the measurement of stress during discrete time periods and in semi lab conditions. Here we describe our efforts to develop a demonstrator system to measure stress in daily life, using continuous sensor input. From the prototypes and relevances obtained from the LVQ methods we found which modalities were most expressive in the classification task. The most prominent modality was found to be Electrocardiogram (ECG) followed by Galvanic Skin Response (GSR) and Respiration (RSP) in terms of performance. Most features contributing to the final relevance matrix of GMLVQ, however, originated from GSR. Another criterion for selecting a suitable modality for measuring affect from physiology in daily life encompasses the obtrusiveness of the sensor technology. In order to maximize acceptance, unobtrusive sensing is preferred to allow the users to move without restrictions (Westerink et al. 2012). On-body sensors might be preferred because they pose no restrictions on range of movement of the user without loosing connection with the sensors.

The modalities explored in Chapter 4 (ECG, RSP, and GSR) can be measured on the body using either a chest strap (for ECG, and RSP) or a sensor on a hand or foot (for GSR). Wearing a chest strap throughout days is uncomfortable, making ECG and RSP unfavorable. Hence our preference goes to GSR sensors. The traditional

position for measuring GSR, on the sole of the feet or palm of the hands, is however impractical as it interferes with grasping and touching (hands) and is subject to variable pressure that influences the measurement (both hands and feet). From an experiment in which we tested suitability of GSR measurement on 16 locations on the body (van Dooren et al. 2012), we concluded that next to the feet and hands, the best location for measuring emotional reactions from GSR is on the wrist. By integrating the GSR sensors in a wrist-worn device, similar to a wrist-watch, we obtain an unobtrusive measurement of GSR (Westerink et al. 2009).

In this section we describe an application based upon stress measurement from skin conductance measured on the wrist. The platform we use for these measurements is based upon our emotion measurement platform (Westerink et al. 2009) in which the GSR sensor applies a small (direct) current over two electrodes attached to the skin. Via measurement of the voltage over a reference resistor, the resistance of the skin can be obtained. The signal is sampled at 2Hz, analyzed in real time and stored in flash memory on the device. Further details on the hardware used can be found in Westerink et al. (2014).

As discussed in van den Broek, van der Zwaag, Healey, Janssen and Westerink (2010) one of the main challenges in real time affective signal processing is to select a suitable means of correcting the signals (or features) for variations amongst users or long term variations within users. To this end, some form of normalization using a reference signal is required. The challenge consists of two choices to be made. Firstly, to select a time period over which the signal is considered a reference signal, and secondly, to select a method to correct the signal using the reference signal. As opted by van den Broek, van der Zwaag, Healey, Janssen and Westerink (2010), we employ a sliding window to define the period over which a reference signal is taken.

The method we apply is chosen based upon two criteria: robustness against outliers, and preferably have a fixed-interval range. Daily life measurements are inherently noisy. Therefore all steps in the signal analysis process should be designed such that they are robust against outliers. Many of the normalization techniques used in laboratory conditions, especially those using minimal or maximal values, such as discussed in van den Broek, van der Zwaag, Healey, Janssen and Westerink (2010), are for that reason not suitable. Because our daily life stress indicator aims at measuring an absolute level of stress (i.e., on a fixed interval), we prefer normalization techniques that return values on a fixed-interval range, such that they can be easily mapped to the fixed-interval indication we are aiming at. The often used z-correction, which was also used in the experiment described in this chapter, does

Figure 7.1: *The Vitality Bracelet with all (green) lights for paced breathing on as well as all (blue) lights indicating the duration.*

not limit the range of its output and is therefore, in its pure form, not preferable. The normalization method we developed is, however, inspired by z-correction, since it also makes use of the distribution of the data. Z-correction basically assumes that the data is normally distributed, and after correction (subtraction of the mean and division by the standard deviation), yields standard normally distributed data (i.e., if the assumption holds). We observed that in everyday life, skin conductance measurements taken in a sliding window seldom follow a normal distribution. Rather than assuming normality, we use the distribution of the measurements in the sliding as a reference signal and use the relative position of the current sample with respect to the reference distribution as a normalized signal.

To this end, as described in de Waele et al. (2009) and Westerink et al. (2014), we calculate the cumulative histogram from the measurements taken in the sliding window, and by applying the cumulative histogram to the current sample, we find the relative position the current sample has with respect to the distribution. By normalizing the histogram (division by the number of samples in the moving window), we ensure the normalized values to be in the range $[0, 1]$. More formally, for the latest sample x_N in a time series $(x_1 \dots x_N)$ we calculate H_x^τ using the following equation:

$$H_x^\tau(x_N) = \frac{1}{\tau}|\{x_n : x_n \leq x_N \wedge N - \tau \leq n < N\}|. \tag{7.1}$$

where τ defines the the length of the sliding window (in number of samples). When the value of H_x^τ exceeds a given threshold T_H (which can be considered a percentile of the values $x_{N-\tau} \dots x_N$ in the window with length τ), the system triggers the detection of a physiological stress response. When τ is chosen to represent a time window of several minutes, the triggers represent instantaneous stress, while a time window in the order of an hour will represent the handful of really

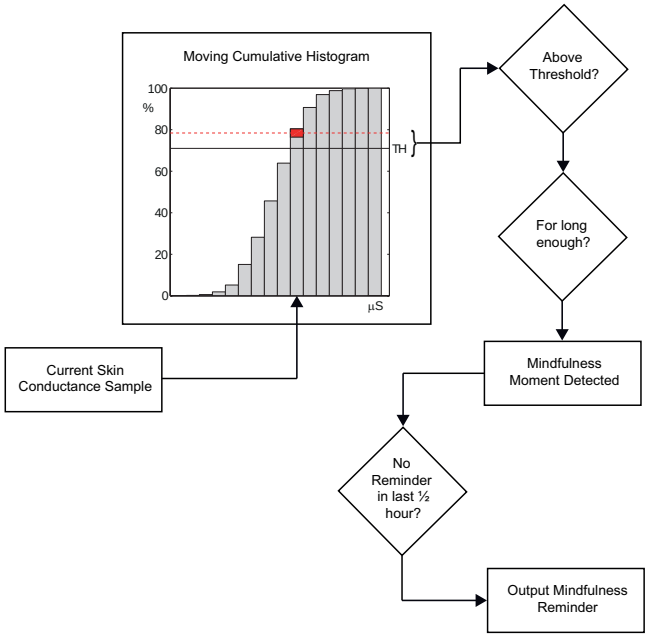

Figure 7.2: *Graphical representation of the mindfulness algorithm used in the Vitality Bracelet.*

stressful events that happen daily during everyday life. In order to improve further robustness of the algorithm, the threshold should be exceeded for at least t samples, thereby defining a stress alert SA as:

$$SA \iff \forall N - t < n \leq N :: x_n > T_H \tag{7.2}$$

In Westerink et al. (2014) the concept of the Vitality Bracelet is presented in which this algorithm is used. The Vitality Bracelet (see Figure 7.1) uses the stress alerts to highlight stressful events, referred to as Mindfulness Moments, to the wearer by means of a gentle vibration. We empirically found mindfulness reminders to lose their effectiveness when being presented in too short succession. In order to avoid indicating a second (possibly related) mindfulness moment shortly after a first has happened, we finally restrict the number of Mindfulness Moments given to at most one in the last 30 minutes. The algorithm for these mindfulness moments is graphically depicted in Figure 7.2.

In order to handle and relief the stress that caused the Vitality Bracelet to trigger

a Mindfulness Moment, the user is suggested to take a moment of reflection and is offered a paced breathing exercise, to be taken at any moment the user prefers. The paced breathing exercise is guided by the Vitality Bracelet by means of a series of lights laid out in a spiral that indicate a preferred breathing pace that gradually reduces the pace over the first 5 minutes. The duration of the exercise is indicated up to 15 minutes. During daily life tests the Vitality Bracelet and the mindfulness algorithm correctly identified various stressful moments such as ”stepping into the dentist’s chair”.

Figure 7.3: *Pictures of famous faces; (left) Mona Lisa, by Leonardo Da Vinci*[1]*; (mid) Bill Clinton*[2]*; (right) Albert Einstein*[3].

[1]source: http://en.wikipedia.org/wiki/File:Mona_Lisa,_by_Leonardo_da_Vinci,_from_C2RMF_retouched.jpg

[2]source: http://www.nieuwsblad.be/article/detail.aspx?articleid=DMF20130119_054

[3]source: http://www.laboiteverte.fr/16-portraits-dalbert-einstein/

7.3 Facial Expressions

Following the example of Sebe and colleagues (Sebe 2005, Sebe 2006), we applied the classifiers trained for facial expression recognition, as described in Chapter 5 to several well known faces. In their original attempt, Sebe et al., applied a system developed at the University of Illinois (Cohen et al. 2003) to the Mona Lisa. This system uses a wire-frame or facial mesh from which deformations are extracted and used as features, similar to the action units of Ekman's Facial Action Coding System (Ekman et al. 2002). "It concluded that the subject was 83% happy, 9% disgusted, 6% fearful and 2% angry, New Scientist magazine was told." (Sebe 2005)

To that end, we searched for a high quality picture of the Mona Lisa painting by Leonardo da Vinci, Bill Clinton during his statement on the Lewinsky case, and Albert Einstein (see Figure 7.3). We applied the preprocessing as described in Section 5.3 and applied both 6 and 7-class RSLVQ classifiers that were trained on the Cohn-Kanade database.

The class-wise posterior probabilities as defined by RSLVQ ($P(S|\xi^\mu)$ of Equation (2.12)) allow for a similar soft characterization of the facial expressions present, as done by Sebe (2005). These results are given in Table 7.1. It can be observed that the 7-class classifier assigns the Neutral category most predominantly ($> 95\%$), with small contributions of Sadness, Disgust, Anger and Happiness. The 6-class classifier forces the choice to the 6 emotions of which the Sadness is most prominent (38%), followed by Disgust (25%) and smaller contributions of the other emotions. Compared to the reported class assignment (Sebe 2005), our classifier assigns far less Happiness, but a mixture of emotions or rather assigns Neutral.

The results for Bill Clinton and Albert Einstein can be found in Tables 7.2 and 7.3. They indicate that Bill Clinton also shows, next to a relatively neutral face, a mix of Sadness and Surprise. The picture of Albert Einstein shows predominantly Surprise. Given the high accuracies obtained by our classifiers on the Cohn-Kanade database, we can say with high certainty that, out of the set of emotions trained, these are the emotions represented in these iconic faces.

Table 7.1: *Probabilities of assigning class labels to the Mona Lisa image by RSLVQ.*

Class label	7-class classification	6-class classification
Anger	0.92%	10.06%
Disgust	1.11%	24.57%
Fear	0.06%	6.82%
Happiness/Joy	0.98%	8.18%
Neutral	95.49%	-
Sadness	1.40%	38.40%
Surprise	0.03%	11.96%

Table 7.2: *Probabilities of assigning class labels to the Bill Clinton image by RSLVQ.*

Class label	7-class classification	6-class classification
Anger	0.39%	1.09%
Disgust	1.43%	4.27%
Fear	2.59%	7.30%
Happiness/Joy	18.71%	11.74%
Neutral	37.79%	-
Sadness	28.73%	38.64%
Surprise	10.36%	36.96%

Table 7.3: *Probabilities of assigning class labels to the Albert Einstein image by RSLVQ.*

Class label	7-class classification	6-class classification
Anger	0.01%	0.03%
Disgust	0.03%	0.14%
Fear	0.16%	0.48%
Happiness/Joy	0.18%	0.45%
Neutral	0.05%	-
Sadness	0.06%	0.19%
Surprise	99.51%	98.71%

7.4 Empathic Photo-Frame

The mechanism of using appraisal dimensions to link pictures with their emotional experience was implemented in a prototype named the Empathic Photo Frame. Traditional Photo Frames, including the popular electronic variant, show a (playlist of) picture(s) that has been preselected and configured. Such photo frames do not adapt to the state of the user. The Empathic Photo Frame enables an automated way of adapting a dynamic playlist to the emotional state of a user. Key to the application is a matching dimensional model of emotions to which both the photos and emotional state representation can be mapped. This could be the 2-dimensional circumplex model by Russell and Mehrabian (1977), the 8-dimensional model by Cochrane (2009), other dimensional models (Mehrabian and Russell 1974, Watson and Tellegen 1985, Russell 2003, Lövheim 2012), or dimensional appraisal models (Scherer 2001, Smith and Lazarus 1990), such as the CPM, which was used in this prototype.

After selecting a desired state represented in the dimensional model, the closest photos to that state can be selected and added to the top of the play list to be shown to the user. Based upon user feedback, the internal representations can be updated during real time use, in order to enable adaptation to specific users and to improve future performance. A Matlab (R2013a) implementation was created to emulate the behavior of a photo frame and perform the photo selection and real time adaptation. Figure 7.4. Three parts of the system will be treated in further detail in the following sections:

- The mapping of stimuli and emotional states to the dimensional model
- The selection of a desired state, emotional journey selection, and subsequent photo selection and playlist adaptation
- Realtime adaptation and optimization

7.4.1 Mapping to a dimensional model

Both photos and emotional states need to be represented in the dimensional model because distances in the space spanned by the dimensional model are used for stimulus selection. The mapping of emotional states to the dimensional model is a one-time action, for which the prototypes as found in the classification task described in Section 6.5.2 can be used. For this prototype, the photos used were rated using the CPM appraisal questionnaire (Table 6.1), but for practical application, the

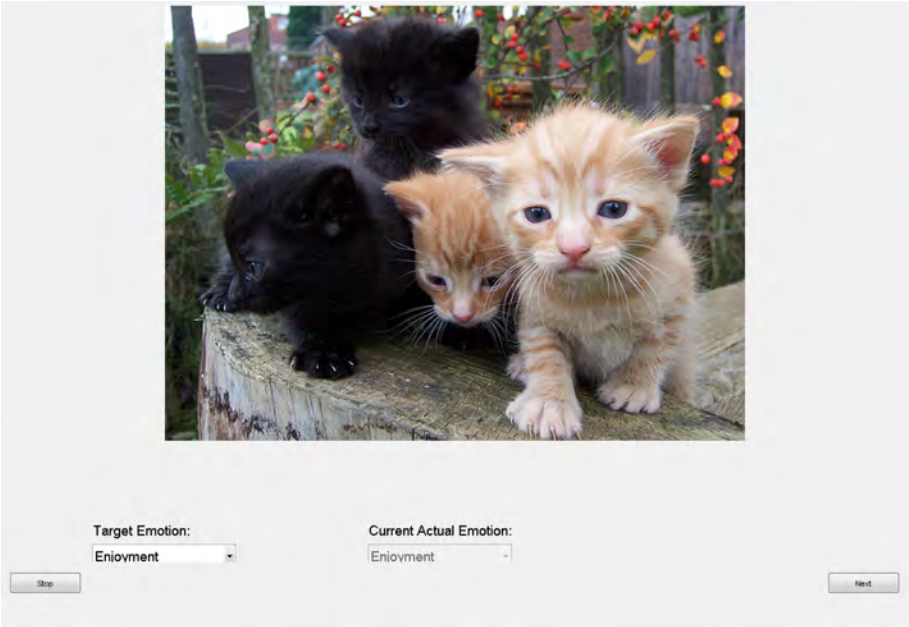

Figure 7.4: *Graphical User Interface of the Empathic Photo frame showing a picture that is targeting the selected emotion (lower left) of Enjoyment.*

mapping of stimuli to the appraisal space should be performed in an automated fashion as personal photo collections are dynamical and growing larger and larger in recent years. Since we humans are able to interpret pictures on different affective dimensions, in principal an automated system should be able to (at least to certain extent) perform a similar reasoning. We foresee three alternative sources that can be used as input for this mapping:

- Affective Image Classification
- Affective Text Classification
- Affective Reaction Classification

Affective Image Classification

Much information is embedded in the pixels of which the pictures are composed of. Features describing the composition of photos in terms of colors, textures, but potentially also dedicated detection of objects or people can be used to position pictures in the affective space spanned by the dimensional model. Attempts in the

field of Affective Image Classification (Machajdik and Hanbury 2010, Valdez and Mehrabian 1994) indicate that this should be feasible using machine learning techniques.

Affective Text Classification

Next to the content of photos themselves, more and more metadata becomes available as pictures are shared on social media and commented upon by the owner of the picture as well as others. Comments can occur in the form of tags (sets of single word descriptors), or in the form of longer descriptions using natural language. Through various Natural Language Processing (NLP) techniques the textual information can be mined for affective information which can be used in automated classifiers or regressors in affective text recognition (Binali et al. 2010, Scherer 2005, Keshtkar and Inkpen 2010, Liu et al. 2013). Potentially the textual information can be enriched with other meta information such as the location and time a picture was taken.

Affective Reaction Classification

A third potential method is the positioning of pictures in the dimensional affective space based upon the affective reaction people show when watching a photo. From measurement of physiological, auditory, facial or postural reactions while people are watching stimuli, features could be identified that can be mapped to the affective dimensions. Chapters 4 and 5 describe the possibility of using such signals to classify emotions, which suggests the potential of these signals to be mapped to other affective dimensions as well.

7.4.2 Desired state selection and playlist adaptation

With the stimuli mapped to the dimensional affective model, depicted in Figure 7.5, a distance measure applied to the space spanned by the dimensional model can be used to select stimuli that are close to a chosen desired state in the affective space. By viewing these photos, the emotions felt by the user are expected to be close to the desired state. In this way the user can direct his emotional feelings towards a desired state, or along a desired emotional path.

Desired state selection

In principal, a user could freely indicate a position in the affective space as desired affective state. Affective dimensions are, however, not always easily interpretable

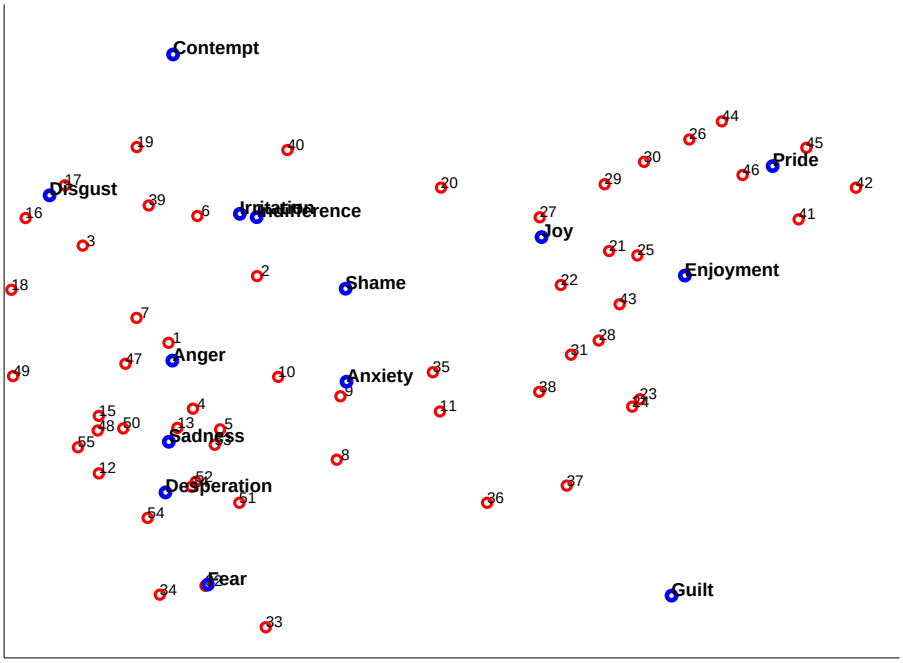

Figure 7.5: *Stimuli (numbered red circles) and emotions (blue circles) mapped to the space spanned by the dimensional affective model.*

and choosing a position in a high dimensional space (e.g., 14 dimensions for the CPM) is far from trivial. Therefore we use the mapping of (words describing) emotions in the affective space as a selection of positions from which the user can choose. By choosing one of 15 emotions, the user sets his desired state which will be used for stimulus selection through playlist adaptation.

Playlist adaptation

Having set a desired state, stimuli can be ordered by ascending distance towards the desired state, as depicted in Figure 7.6. In order to avoid the same photo to be chosen each time a certain emotion is selected as desired state, some randomness should be built in. We used the inverse rank of the top 10 closest stimuli as weights used to throw a weighted 10-sided dice. Formally, consider the ordered set S' that contains the stimuli by ascending distance towards the desired state (hence, s'_1 is the picture closest to the desired state). The next stimulus s'_i to be shown is drawn from

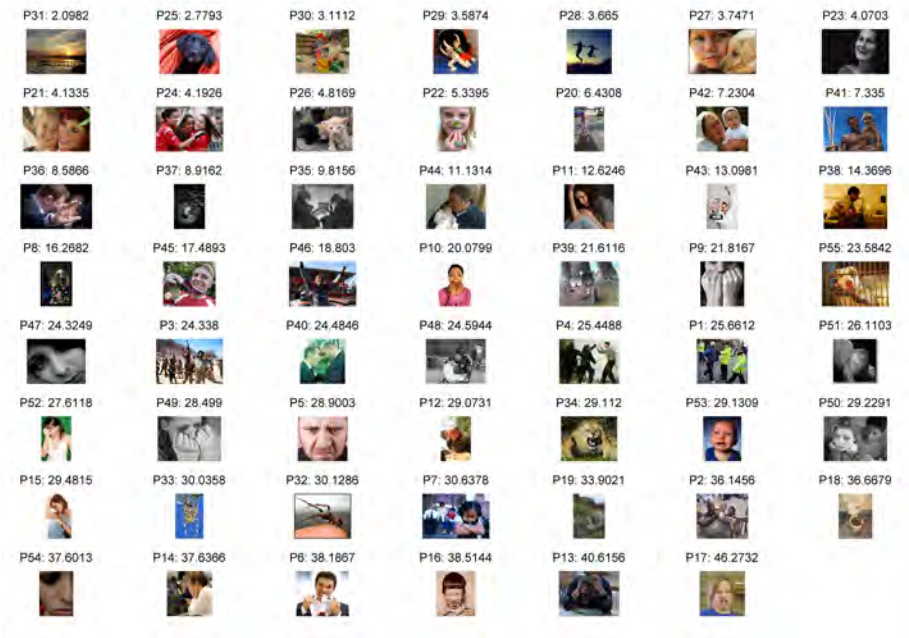

Figure 7.6: *Photos ordered by distance towards the selected desired state.*

the following probability density function:

$$p(s'_i) = \begin{cases} \frac{N_{top}-i+1}{N_{top}*(N_{top}+1)/2} & \text{if } 1 \leq i \leq N_{top} \\ 0 & \text{otherwise} \end{cases} \tag{7.3}$$

where N_{top} indicates the number of closest stimuli that are considered for selection, which we chose to be 10. A potential extension could be the inclusion of the distances in the weighting function that forms the probability density function.

7.4.3 Realtime adaptation and optimization

During the use of the empathic photo frame, the system adapts to the user and learns about their emotional interpretation of the pictures. This is empowered by the use of user-feedback.

User Feedback

We foresee two types of feedback that can be used for this purpose. The first is a manual entry of the emotion felt after each photo the user considers wrongly selec-

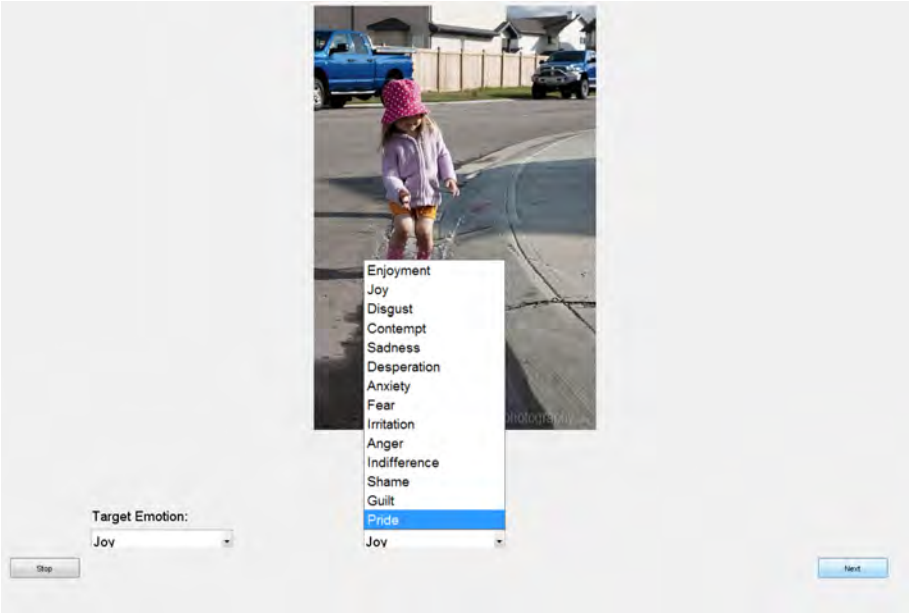

Figure 7.7: *Feedback provided by the user through the selection box (mid low), indicating that the picture triggered the emotion of Pride rather than Joy.*

ted (given the desired state selected), which is shown in Figure 7.7. Alternatively, the emotional reaction of the user watching the picture could be measured through, e.g., facial expressions or physiology. This approach would require a reliable mapping of measurement data to the affective space. There are various indications that this might be feasible for individual affective dimensions (Aue et al. 2007, Grandjean and Scherer 2008, Bradley et al. 1993, Smith 1989, van Reekum et al. 2004), but has not yet been shown fully as to date.

System update

As a first step we integrated physiological measurements of arousal, measured through GSR. Rather than mapping the arousal measurement to the affective space, we used it to scale the size of the update. Similar to the learning rate in LVQ we use the normalized arousal measurement as adaptive learning rate during realtime optimization. Whenever the user provides feedback by indicating which emotion was induced by the picture shown, the GSR value is compared to the window of most recent GSR samples using the cumulative histogram, and the quantile corresponding to the current sample is calculated, which results in a normalized value in the

interval $[0, 1]$, using the method described in de Waele et al. (2009). Formally, if we name the GSR signal $x(t)$ at time t, the normalized GSR value, using the cumulative histogram H_τ over τ seconds is defined as:

$$H_\tau(t) = \frac{1}{\tau}|\{x(u) : x(u) \leq x(t) \wedge t - \tau < u < t\}|, \tag{7.4}$$

where $|\{.\}|$ indicates the number of elements in the set. Following popular LVQ variants, the closest correct and incorrect prototypes are updated by an attractive and repellant force, respectively. The normalized Skin Conductance value moderates the update size as multiplicative factor on top of the default learning rate. We used the RSLVQ scheme (Equation (2.13)) to define the updates such that the update scheme becomes:

$$\begin{aligned} w_J^{\mu+1} &= w_J^\mu + \frac{\eta_{arousal}\eta}{v_{soft}} \left(P_J(T|\xi^\mu) - P(T|\xi^\mu)\right)(\xi^\mu - w_J^\mu), \\ w_K^{\mu+1} &= w_K^\mu - \eta_{arousal}\eta P(T|\xi^\mu)(\xi^\mu - w_K^\mu), \end{aligned} \tag{7.5}$$

where $\eta_{arousal} = H_{10}(t)$, thus using an interval of 10 seconds for the cumulative histogram. w_J and w_K refer in this case to the prototypes representing the user-feedback emotion, and the targeted (incorrect) emotion, respectively. In addition, we allow pictures to move into the direction of the user-feedback emotion in case of corrective feedback. The picture is in that case also considered a closest incorrect prototype and is updated with the update described for w_K in Equation (7.5), however with (ten-fold) larger learning rate. In this way we allow the prototypes to move slowly to accommodate for generic differences in emotional experience between users, and pictures to move more freely to correct for user specific emotional interpretation of individual pictures. Consider for example a picture of a small child playing. This can trigger different emotions in a generic audience versus the parents of the child. Figure 7.8 shows a system of emotions and pictures in the affective space, and how it evolved from the initial situation (Figure 7.5) through user feedback.

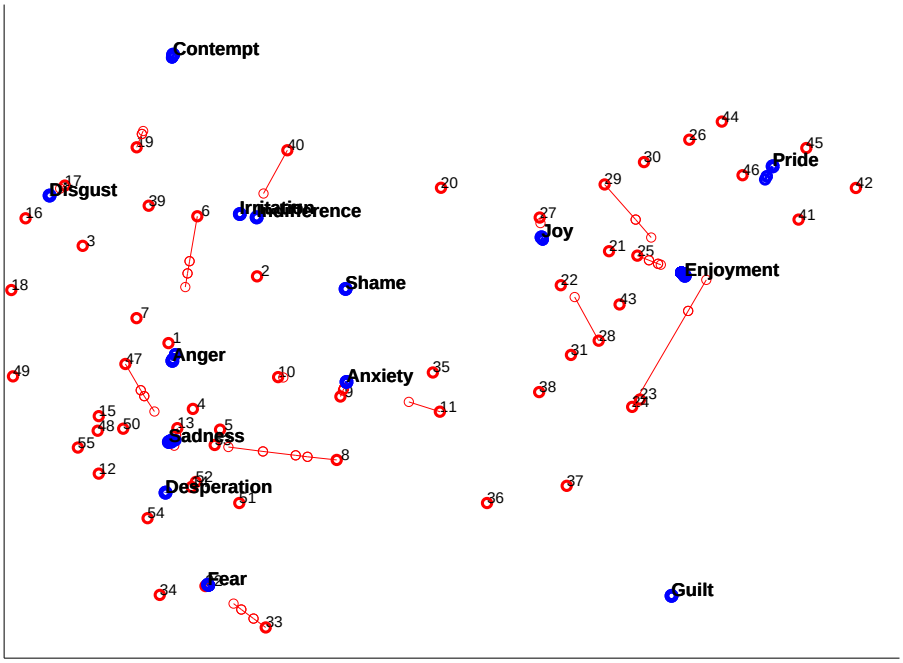

Figure 7.8: *Stimuli and emotions mapped to the space spanned by the dimensional affective model after several update steps using the user's feedback indicating movement of representations of both emotions and pictures in the affective space; compare to the initial state depicted in Figure 7.5.*

Hoofdstuk 8

SUMMARY

Summarizing, this thesis focusses on advancing the knowledge of LVQ methods, in particular in the application to domain of affective computing. In this thesis various LVQ variants were studied in a controlled environment characterized by high dimensional isotropic clusters. This data model yields a simplified representation of reality that allows studying learning dynamics in detail. The findings demonstrated that while LVQ 2.1 can obtain optimal asymptotic performance it suffers from stability issues due to divergent behavior of the prototypes. The introduction of a window, in LFM-W, shows convergence for well chosen window sizes, but introduced large dependency on parameter optimization, with respect to reaching optimal performance. While GLVQ suffered from both divergence and parameter dependency, RSLVQ combined the strengths of both methods and, as the name suggests, is more robust to the parameter settings. It required only limited parameter optimization to reach close to optimal asymptotic performance. Optimization of the window size and softness parameter of RSLVQ, however was critical for efficient learning.

Affective computing comprises the application of computers to the study of emotions, moods, stress and other affective phenomena. Although affect has been studied for a long time in psychology, the research field of affective computing is still in its relative infancy and poses interesting challenges to computer science and in particular machine learning. In order to research the performance of LVQ variants outside of the controlled simulation environment, we applied various LVQ techniques to various classification tasks in the affective domain. To the best of our knowledge, this work comprises a first application of LVQ to the particular affective challenges at hand. Applied to these three real world scenarios, RSLVQ showed its robustness and very competitive predictive performance as affective classifier, competing well with GMLVQ and linear SVM. To that end, three datasets were collected using different input modalities, that approached affective computing from three perspectives:

Bodily A large-scale study was performed in which skin conductance, respiration and electrocardiogram were measured in semi-controlled conditions. A large variety of features was derived from each of these modalities. Classifiers were built using uni- and multi-modal input. Using LVQ techniques, we obtained up to 86.7% accuracy and AUC of 0.95 in the two-class classification task to separate stress from relaxation. Relevance learning was used to identify the most informative features, indicating that most information was embedded in the cardiac signals primarily in time-domain HRV measures. In addition to commonly used features, we also explored various novel features, of which the very-high frequency band of the power spectrum was found to be a very relevant addition. Best performing algorithms were GMLVQ, RSLVQ and linear SVM, of which the performance differences were within percentage points.

Facial A benchmark dataset, the Cohn-Kanade database, was used to build classifiers for facial expression recognition, based upon LBP features that reflect local textures from still images. The feature space was characterized by its large dimensionality. Linear SVM performed best with accuracies up to 94.5% for 6-class and 93.2% for 7-class classification, closely followed by RSLVQ. Implicit relevances obtained from RSLVQ provided insight into the most prominent features, which originated, primarily, from the mouth region and eye regions. The specific LBP features that were found most influential within these regions confirmed that mouth opening/closing is a primary differentiator picked up by the classifier.

Cognitive The cognitive processes involved in the emotional interpretation of stimuli is termed appraisal. Several appraisal models have been coined that define several dimensions of emotional appraisal. We set up a study in which emotions were elicited and accompanying appraisal values determined. In a 5-class classification task, RSLVQ and SVM performed similarly well with accuracies up to 54.5% and Cohen's Kappa of 0.42. Compared to the earlier applied technique of class conditional means, this showed a gain of 5 percentage point in accuracy and 0.05 in Kappa. Applied to this noisy dataset GMLVQ performed relatively poorly at a level similar to that of GLVQ.

Observations made over the three experiments include that GLVQ and GRLVQ suffered from stability issues similar to the simulated environment. RSLVQ and linear SVM performed similarly well at performance levels close to or beyond the state of art. GMLVQ performed poorly in a very noisy environment, but very well in the relatively easier classification tasks. The computational costs involved in relevance learning, however did not allow for the application of GMLVQ to the facial

emotion dataset.

For each of the three affective domains we outlined potential applications and demonstrators. These include the *Vitality Bracelet*, which monitors daily stress levels through skin conductance measurements, triggers in case of high stress and provides a paced breathing exercise to let the user calm down; the facial expression recognition was applied to various iconic faces to reveal the emotions displayed in the Mona Lisa, by Bill Clinton and Albert Einstein; and usage of the cognitive appraisal model was explored in the *Empathic Photo Frame* that displays pictures to match a desired emotion while using representations of both pictures and emotions in the space spanned by the appraisal dimensions.

8.1 Outlook

The following topics have been identified for future work:

- Although in our experiments we made steps to move the affective research from the lab to more daily life situations, the conditions in our experiments were still semi-controlled. It would be interesting to see how the classifiers trained and the results found translate to even less controlled, daily life situations.
- Very high dimensional spaces such as the dataset for facial emotion recognition pose their challenges for the more computational expensive relevance learning LVQ. Despite our efforts in optimization of the implementation of GMLVQ and GRLVQ these techniques took significant time for such very high dimensional classification problems. Further code optimization or the use of more powerful computers would enable the application of relevance learning in LVQ on very high dimensional datasets.
- We have outlined demonstrator systems for the three affective subdomains researched in this thesis. Further validation and optimization of these systems in a daily life setting could bring these applications closer to productization.

In sum this thesis provides useful results for further investigations of the application of LVQ methods to the affective domain and brings us one step closer to the application of affective computing in daily life.

Publications

The following lists a selection of publications and patents directly related to the work presented in this thesis. A complete list of publications is available upon request.

Journal publications published

de Vries, J. J. G., Lemmens, P. M. C., Brokken, D., Pauws, S. C. and Biehl, M.: in press, Towards emotion classification using appraisal modeling, *International Journal of Synthetic Emotions* .

de Vries, J. J. G., Pauws, S. C. and Biehl, M.: in press, Insightful stress detection from physiology modalities using learning vector quantization, *Neurocomputation* .

Janssen, J. H., IJsselsteijn, W. A., Westerink, J. H. D. M., Tacken, P. and de Vries, G.-J.: 2013, The tell-tale heart: Perceived emotional intensity of heartbeats, *International Journal of Synthetic Emotions* **4**(1), 65–91.

Janssen, J. H., Tacken, P., de Vries, J. J. G., van den Broek, E. L., Westerink, J. H. D. M., Haselager, P. and IJsselsteijn, W. A.: 2013, Machines outperform lay persons in recognizing emotions elicited by autobiographical recollection, *Human Computer Interaction* **28**(6), 479–517.

van Dooren, M., de Vries, J. J. G. and Janssen, J. H.: 2012, Emotional sweating across the body: Comparing 16 different skin conductance measurement locations, *Physiology & Behavior* **106**(2), 298–304.

Witoelar, A. W., Ghosh, A., de Vries, J. J. G., Hammer, B. and Biehl, M.: 2011, Window-based example selection in learning vector quantization, *Neural Computation* **22**(11), 2924–2961.

Journal publications submitted

de Vries, J. J. G., Pauws, S. C. and Biehl, M.: submitted 2014, Facial expression recognition using learning vector quantization, *Pattern Recognition* .

Conference contributions

Berezhnoy, I. J., de Vries, G.-J., Weysen, T., Dimov, J. and Garcia-Molina, G.: 2012, Towards unobtrusive automated sleep stage classification - polysomnography using electrodes on the face, *Proceedings of the fifth International Conference on Health Informatics (HEALTHINF)*, pp. 487–492.

de Vries, G.-J. and Biehl, M.: 2009, Analysis of robust soft learning vector quantization and an application to facial expression recognition, *in* M. Biehl, B. Hammer, S. Hochreiter, S. C. Kremer and T. Villmann (eds), *Similarity-based learning on structures*, Dagstuhl Seminar Proceedings, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany, Dagstuhl, Germany.

de Vries, G.-J., Lemmens, P. and Brokken, D.: 2009, Same or different? recollection of or empathizing with an emotional event from the perspective of appraisal models, *3rd International Conference on Affective Computing and Intelligent Interaction and Workshops (ACII 2009)*, pp. 1–6.

de Vries, G.-J. and van der Zwaag, M. D.: 2010, Enhanced method for robust mood extraction from skin conductance, *Proceedings of the third International Conference on Bio-inspired Systems and Signal Processing (BIOSIGNALS)*, Valencia, Spain, pp. 139–144.

de Vries, J. J. G., van Dooren, M., van Beek, W. H. M., Dijk, E. O., Ouwerkerk, M. and Westerink, J. H. D. M.: 2012, Deriving stress from peripheral physiology, *Proceedings of the 33rd International Conference of the Stress and Anxiety Research Society (STAR)*, Palma de Mallorca, Spain, p. 108.

de Waele, S., de Vries, G.-J. and Jäger, M.: 2009, Experiences with adaptive statistical models for biosignals in daily life, *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pp. 1–6.

Lemmens, P., Crompvoets, F., Brokken, D., van den Eerenbeemd, J. and de Vries, G.-J.: 2009, A body-conforming tactile jacket to enrich movie viewing, *EuroHaptics conference, 2009 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2009. Third Joint*, pp. 7–12.

Lemmens, P. M. C., Brokken, D. and de Vries, G.-J.: 2010, Tactile experiences, *in* A. Nijholt, E. O. Dijk and S. Luitjens (eds), *The EuroHaptics 2010 Special Symposium: Haptic and Audio-Visual Stimuli: Enhancing Experiences and Interaction*, Amsterdam, the Netherlands, pp. 11–17.

Westerink, J., Ouwerkerk, M., de Vries, G.-J., de Waele, S., van den Eerenbeemd, J. and van Boven, M.: 2009, Emotion measurement platform for daily life situations, *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pp. 1–6.

Westerink, J., van Boxtel, A., IJsselsteijn, W., Janssen, J., Ouwerkerk, M., Overbeek, T., de Vries, G.-J., Slovak, P., van der Zwaag, M. and Fitzpatrick, G.: 2012, Unobtrusive emotion sensing in everyday life, *in* A. J. Spink, F. Grieco, O. E. Krips, L. W. S. Looijens, L. P. J. J. Noldus and P. H. Zimmerman (eds), *Proceedings of Measuring Behavior 2012, 8th International Conference on Methods and Techniques in Behavioral Research*, Utrecht, the Netherlands, p. 332.

Zangouei, F., Babazadeh Gashti, M. A., Höök, K., Tijs, T., de Vries, G.-J. and Westerink, J.: 2010, How to stay in the emotional rollercoaster: Lessons learnt from designing EmRoll, *The sixth Nordic Conference on Human-Computer Interaction (NordiCHI)*, Reykjavik, Iceland, pp. 571–580.

Book chapters

Westerink, J., van Beek, W., Daemen, E., Janssen, J., de Vries, G.-J. and Ouwerkerk, M.: 2014, The vitality bracelet: Bringing balance to your life with psychophysiological measurements, *in* S. H. Fairclough and K. Gilleade (eds), *Advances in Physiological Computing*, Springer London, London, pp. 197–209.

Patents & Patent Applications

Berezhnyy, I. and de Vries, J. J. G.: 2012, Sleep stage classification device with background oscillation emitter. WO2013061185.

Crompvoets, F. M. H., de Vries, J. J. G., Brokken, D., van den Eerenbeemd, J. M. A. and Lemmens, P. M. C.: 2009, System, method and computer program product for indicating stimulation signals to a user. WO2010100588, EP2403467, US8527041.

de Vries, J. J. G. and Ouwerkerk, M.: 2011, Stress-measuring device and method. WO2012140537, US20140031704, EP2696754.

Dijk, E. O., Janssen, J. H., Westerink, J. H. D. M., de Vries, J. J. G. and van Dooren, M.: 2011, Breathing guidance device and method. WO2012117376.

Lashina, T. A. and de Vries, J. J. G.: 2012, Selection of ambient stimuli. WO2013144854.

Newton, P. S., van Doveren, H. F. P. M., Brokken, D., van den Eerenbeemd, J. M. A., Crompvoets, F. M. H., Lemmens, P. M. C. and de Vries, J. J. G.: 2009, Methods and systems for adapting a user environment. WO2010119376, US2012-0041917, EP2419184.

van den Eerenbeemd, J. M. A., Crompvoets, F. M. H., Brokken, D., de Vries, J. J. G. and Lemmens, P. M. C.: 2009, Method and system for processing a physiological signal. WO2010119374, EP2419002, US2012-0046875.

Samenvatting

Dit proefschrift bestudeert verschillende Lerende Vector Quantisatie (LVQ) algoritmen in een gesimuleerde omgeving. Het gebruikte model, bestaande uit hoog dimensionale isotropische centra, maakt een gedetailleerde studie van de dynamica van het leerproces mogelijk. De resultaten geven aan dat LVQ 2.1 asymptotisch optimaal kan presteren ondanks het optreden van stabiliteitsproblemen veroorzaakt door divergerende prototypes. Toevoeging van een venster, in LFM-W, levert convergerende prototypes mits de grootte van het venster goed gekozen wordt. De methode wordt daardoor echter wel meer afhankelijk van parameterkeuze voor het behalen van optimale prestaties. Waar GLVQ last heeft van zowel divergentie als van afhankelijkheid van parameterkeuze, combineert RSLVQ de voordelen en is, zoals de naam suggereert, robuuster met betrekking tot de keuze van parameters. Met beperkte parameteroptimalisatie bereikt RSLVQ nagenoeg optimale asymptotische prestaties. De leersnelheid van RSLVQ is echter wel sterk afhankelijk van de venstergrootte.

Affectieve informatica omvat het toepassen van computers in het bestuderen van emoties, stemmingen, stress en andere affectieve fenomenen. Ondanks dat affect al geruime tijd in de psychologie wordt bestudeerd, staat de affectieve informatica nog relatief in de kinderschoenen en biedt nog altijd interessante uitdagingen aan de informatica, vooral als toepassingsgebied van zelflerende systemen. Naast eerder genoemde gecontroleerde omstandigheden, hebben wij de prestaties van LVQ bestudeerd in toepassingen op verschillende affectieve classificatietaken. Voor zover wij weten, zijn wij de eersten die LVQ hebben toegepast op drie affectieve uitdagingen uit het dagelijks leven. RSLVQ toonde ook hier robuustheid en presteerde goed als affectieve classificator, evenals GMLVQ en SVM. Deze resultaten werden behaald na toepassing op affectieve informatie vanuit de volgende drie perspectieven:

Lijf In een grootschalige studie zijn huidgeleiding, ademhaling en hartactiviteit gemeten in semi-gecontroleerde omstandigheden. Een verscheidenheid aan kenmerken uit deze signalen werd gebruikt als invoer voor de classificatoren die zowel op uni- als op multimodale invoer getraind zijn. De LVQ technieken behaalden tot

87.6% accuratesse en een oppervlakte onder de ROC-curve van 0.95 in het onderscheiden van stress van relaxatie als tweeklasse-probleem. Relevantie-leermethoden gaven aan dat de meeste informatie voor deze taak in het hartsignaal gevonden werd, met name in de tijdsdomein hartslagvariatiekenmerken. Naast veelgebruikte kenmerken, hebben we ook enkele nieuwe bestudeerd, waarvan hartslagvariatie in het zeer-hoge frequentiegebied een zeer waardevolle aanvulling bleek te zijn. De best presterende algoritmen op deze taak waren GMLVQ, RSLVQ en lineaire SVM. Onderling verschilden de prestaties slechts enkele procentpunten.

Gelaat We hebben een classificator voor herkenning van gelaatsuitdrukkingen gemaakt op basis van LBP kenmerken, die lokale textuur van foto's representeren. Lineaire SVM presteerde het best, met een accuratesse van 94.5% in geval van 6 klassen en 93.2% in geval van 7 klassen, op de voet gevolgd door RSLVQ. De impliciete relevanties, gevonden middels RSLVQ, gaven inzicht in de meest belangrijke kenmerken, die voornamelijk van de mond- en oog-regio's bleken te komen. De belangrijkste LBP kenmerken uit deze regio's bevestigen dat het openen/sluiten van de mond de belangrijkste differentiator is die de classificator geleerd heeft.

Brein De cognitieve processen die betrokken zijn bij de emotionele interpretatie van stimuli, worden beschreven in theorieën van cognitieve beoordeling. Er bestaan verschillende cognitieve beoordelingsmodellen die emotionele interpretatie uitsplitsen op verschillende dimensies. In onze studie hebben we emoties opgewekt en de waarden van bijbehorende beoordelingsdimensies gemeten. In de classificatie van 5 emotieklassen presteerden RSLVQ en SVM even goed met een accuratesse van 54.5% en Cohen's Kappa van 0.42. In vergelijking met een eerder gepubliceerde studie is dat een 5 procentpunt hogere accuratesse en een 0.05 hogere Kappa. Door de ruis in deze dataverzameling presteerde GMLVQ minder goed, op het niveau van GLVQ.

Wanneer we de drie experimenten vergelijken, zien we dat GLVQ en GRLVQ last hebben van stabiliteitsproblemen, net als in de gesimuleerde omgeving. RSLVQ en lineaire SVM presteerden op of boven het niveau van bestaande studies. GMLVQ presteerde slecht in een omgeving gekenmerkt door veel ruis, maar zeer goed in de relatief makkelijker classificatietaken. De rekenkosten die gepaard gaan met het leren van relevanties in LVQ waren helaas te groot om GMLVQ toe te passen op de classificatie van gelaatsuitdrukkingen.

Voor ieder van de drie bovengenoemde affectieve domeinen hebben we potentiële toepassingen en prototypes beschreven. De *Vitaliteitsarmband* meet in het

dagelijks leven stressniveaus middels huidgeleiding en geeft een signaal in geval van hoge stress, waarop een begeleide ademhalingsoefening wordt aangeboden met als doel de gebruiker te kalmeren. De gelaatsuitdrukkingsherkenning werd toegepast op verschillende iconische gezichten waaronder de Mona Lisa, Bill Clinton en Albert Einstein. De *Empatische Fotolijst* laat foto's zien die passen bij een gewenste emotie en gebruikt daarbij representaties van de foto's, alsmede emoties in de wiskundige ruimte, die opgespannen wordt door de affectieve beoordelingsdimensies.

Bibliography

Ahmed, F.: 2012, Gradient directional pattern: A robust feature descriptor for facial expression recognition, *Electronics Letters* **48**(19), 1203–1204.

Ahsan, T., Jabid, T. and Chong, U.-P.: 2013, Facial expression recognition using local transitional pattern on gabor filtered facial images, *IETE Technical Review* **30**(1), 47–52.

Ali, H. B., Powers, D. M. W., Leibbrandt, R. and Lewis, T.: 2011, Comparison of region based and weighted principal component analysis and locally salient ICA in terms of facial expression recognition, *in* J. Kacprzyk and R. Lee (eds), *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing 2011*, Vol. 368, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 81–89.

Allen, J. J. B., Chambers, A. S. and Towers, D. N.: 2007, The many metrics of cardiac chronotropy: a pragmatic primer and a brief comparison of metrics, *Biological psychology* **74**(2), 243–262.

Aue, T., Flykt, A. and Scherer, K. R.: 2007, First evidence for differential and sequential efferent effects of stimulus relevance and goal conduciveness appraisal, *Biological Psychology* **74**(3), 347–357.

Backé, E.-M., Seidler, A., Latza, U., Rossnagel, K. and Schumann, B.: 2012, The role of psychosocial stress at work for the development of cardiovascular diseases: a systematic review, *International archives of occupational and environmental health* **85**(1), 67–79.

Baldaro, B., Rossi, N., Caterina, R., Codispoti, M., Balsamo, A. and Trombini, G.: 2003, Deficit in the discrimination of nonverbal emotions in children with obesity and their mothers, *International journal of obesity and related metabolic*

disorders: journal of the International Association for the Study of Obesity **27**(2), 191–195.

Barkai, N., Seung, H. S. and Sompolinsky, H.: 1993, Scaling laws in learning of classification tasks, *Phys. Rev. Lett. 70* **70**, 3167–3170.

Bengio, Y.: 2000, Gradient-based optimization of hyperparameters, *Neural Comput.* **12**(8), 1889–1900.

Berntson, G. G. and Cacioppo, J. T.: 2004, Heart rate variability: Stress and psychiatric conditions, *in* M. Malik and A. J. Camm (eds), *Dynamic Electrocardiography*, Blackwell Publishing, pp. 57–64.

Biehl, M.: 1994, An exactly solvable model of unsupervised learning, *Europhysics Letters* **25**(5), 391–396.

Biehl, M. and Caticha, N.: 2003, The statistical mechanics of on-line learning and generalization, *The handbook of brain theory and neural networks* pp. 1095–1098.

Biehl, M., Freking, A., Ghosh, A. and Reents, G.: 2004, A theoretical framework for analysing the dynamics of LVQ: A statistical physics approach, *Technical Report 2004-9-02, Mathematics and Computing Science, University Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands, December 2004, available from* `http://www.cs.rug.nl/~biehl`.

Biehl, M., Ghosh, A. and Hammer, B.: 2007, Dynamics and generalization ability of LVQ algorithms, *J. Mach. Learning Res.* **8**, 323–360.

Biehl, M. and Mietzner, A.: 1993, Statistical mechanics of unsupervised learning, *Europhysics Letters* **27**, 421–426.

Biehl, M. and Schwarze, H.: 1993, Learning drifting concepts with neural networks, *Journal of Physics A: Mathematical and General* **26**(11), 2651–2665.

Binali, H., Wu, C. and Potdar, V.: 2010, Computational approaches for emotion detection in text, *Digital Ecosystems and Technologies DEST 2010 4th IEEE International Conference on* **37**(5), 172–177.

Boucsein, W.: 1992, *Electrodermal activity*, Plenum Press.

Boucsein, W.: 2012, *Electrodermal Activity*, Springer.

Bouma, A., Mulder, J. and Lindeboom, L.: 1996, *Neuropsychologische diagnostiek: Handboek*, Swets & Zeitlinger, Lisse.

Bradley, M. and Lang, P.: 1994, Measuring emotion: the self-assessment manikin and the semantic differential, *Journal of Behavioral Therapy and Experimental Psychiatry* **25**, 49–59.

Bradley, M. M., Lang, P. J. and Cuthbert, B. N.: 1993, Emotion, novelty, and the startle reflex: habituation in humans, *Behavioral neuroscience* **107**(6), 970–980.

Broekens, J.: 2012, In defense of dominance: PAD usage in computational representations of affect, *International Journal of Synthetic Emotions* **3**(1), 33–42.

Cannon, W. B.: 1967, *The Wisdom of the Body*, W. W. Norton, The Norton Library.

Chanel, G., Kierkels, J. J. M., Soleymani, M. and Pun, T.: 2009, Short-term emotion assessment in a recall paradigm, *International Journal of Human-Computer Studies* **67**(8), 607–627.

Chang, C.-C. and Lin, C.-J.: 2011, LIBSVM: A library for support vector machines, *ACM Transactions on Intelligent Systems and Technology* **2**, 27:1–27:27. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

Choi, J. and Gutierrez-Osuna, R.: 2009, Using heart rate monitors to detect mental stress, *Sixth International Workshop on Wearable and Implantable Body Sensor Networks, 2009. BSN 2009*, pp. 219–223.

Cochrane, T.: 2009, Eight dimensions for the emotions, *Social Science Information* **48**(3), 379–420.

Cohen, I., Sebe, N., Garg, A., Chen, L. S. and Huang, T. S.: 2003, Facial expression recognition from video sequences: Temporal and static modeling, *Computer Vision and Image Understanding* **91**(1-2), 160–187.

Cortez, P. and Embrechts, M. J.: 2013, Using sensitivity analysis and visualization techniques to open black box data mining models, *Information Sciences* **225**, 1–17.

Cottrell, G. W. and Metcalfe, J.: 1991, EMPATH: Face, emotion, and gender recognition using holons, *in* R. P. Lippmann, J. E. Moody and D. S. Touretzky (eds), *Advances in Neural Information Processing Systems*, NIPS Proceedings Series, Morgan Kaufmann, San Mateo, CA, pp. 564–571.

Cox, T.: 1993, *Stress research and stress management: putting theory to work*, Health and Safety Executive.

Darwin, C.: 1872, *On the Expression of the Emotions in Man and Animals*, John Murray.

Datcu, D. and Rothkrantz, L. J. M.: 2005, Facial expression recognition with relevance vector machines, *IEEE International Conference on Multimedia and Expo, 2005. ICME 2005*, pp. 193–196.

Dawson, M., Schell, A. M. and Filion, D. L.: 2000, The electrodermal system, *in* J. T. Cacioppo, L. G. Tassinary and G. G. Berntson (eds), *Handbook of Psychophsiology*, Vol. 2nd, Cambridge University Press, pp. 200–223.

de Vries, G.-J., Lemmens, P. and Brokken, D.: 2009, Same or different? recollection of or empathizing with an emotional event from the perspective of appraisal models, *3rd International Conference on Affective Computing and Intelligent Interaction and Workshops (ACII 2009)*, pp. 1–6.

de Vries, G.-J. and van der Zwaag, M. D.: 2010, Enhanced method for robust mood extraction from skin conductance, *Proceedings of the third International Conference on Bio-inspired Systems and Signal Processing (BIOSIGNALS)*, Valencia, Spain, pp. 139–144.

de Vries, J. J. G., Lemmens, P. M. C., Brokken, D., Pauws, S. C. and Biehl, M.: in press, Towards emotion classification using appraisal modeling, *International Journal of Synthetic Emotions* .

de Vries, J. J. G., Pauws, S. C. and Biehl, M.: in press, Insightful stress detection from physiology modalities using learning vector quantization, *Neurocomputation* .

de Vries, J. J. G., Pauws, S. C. and Biehl, M.: submitted 2014, Facial expression recognition using learning vector quantization, *Pattern Recognition* .

de Waele, S., de Vries, G.-J. and Jäger, M.: 2009, Experiences with adaptive statistical models for biosignals in daily life, *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pp. 1–6.

Dempster, T. and Vernon, D.: 2009, Identifying indices of learning for alpha neurofeedback training, *Applied Psychophysiology and Biofeedback* **34**(4), 309–318.

Douglas-Cowie, E., Cox, C. and et al.: 2006, HUMAINE d5f deliverable, `http://emotion-research.net/download/pilot-db/`.

Dubreuil, B.: 2010, *Human Evolution and the Origins of Hierarchies: The State of Nature*, Cambridge University Press, New York.

Duda, R. O., Hart, P. E. and Stork, D. G.: 2000, *Pattern Classification (2nd Edition)*, Wiley-Interscience.

Ekman, P.: 1972, Universals and cultural differences in facial expressions of emotion, *in* J. Cole (ed.), *Nebraska Symposium on Motivation*, University of Nebraska Press, Lincoln, pp. 207–283.

Ekman, P.: 1979, About brows: Emotional and conversational signals, *in* M. von Cranach, K. Foppa, W. Lepenies and D. Ploog (eds), *Human Ethology: Claims and limits of a new discipline*, Cambridge University Press.

Ekman, P., Friesen, W. V. and Hager, J. C.: 2002, *Facial Action Coding System [E-book]*, Research Nexus, Salt Lake City, UT.

Engel, A. and van den Broeck, C.: 2001, *The Statistical Mechanics of Learning*, Cambridge University Press, Cambridge, UK.

Eriksen, B. A. and Eriksen, C. W.: 1974, Effects of noise letters upon the identification of a target letter in a nonsearch task, *Perception & Psychophysics* **16**(1), 143–149.

Essa, I. A. and Pentland, A. P.: 1995, Facial expression recognition using a dynamic model and motion energy, *in* N. Sebe, M. S. Lew and T. S. Huang (eds), *ICCV'95 Fifth International Conference on Computer Vision*, IEEE Computer Society Press, Cambridge, MA, pp. 360–367.

Essa, I. A. and Pentland, A. P.: 1997, Coding, analysis, interpretation, and recognition of facial expressions, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(7), 757–763.

Fazli, S., Afrouzian, R. and Seyedarabi, H.: 2009, High- performance facial expression recognition using gabor filter and probabilistic neural network, *IEEE International Conference on Intelligent Computing and Intelligent Systems, 2009. ICIS 2009*, Vol. 4, pp. 93–96.

Freeman, S.: 2005, *Biological Science*, 2nd edn, Pearson Prentice Hall, Upper Saddle River, NJ.

Frijda, N. H.: 1987, Emotion, cognitive structure, and action tendency, *Cognition & Emotion* **1**(2), 115–143.

Gale, A. and Edwards, J. A.: 1983, *Physiological Correlates of Human Behaviour: Individual differences and psychopathology*, Academic Press.

Ghosh, A., Biehl, M. and Hammer, B.: 2006, Performance analysis of LVQ algorithms: a statistical physics approach, *Neural Networks* **19**, 817–829.

Giakoumis, D., Tzovaras, D. and Hassapis, G.: 2013, Subject-dependent biosignal features for increased accuracy in psychological stress detection, *International Journal of Human-Computer Studies* **71**(4), 425–439.

Google: 2014, Google glass, http://www.google.com/glass/start/.

Grandjean, D. and Scherer, K. R.: 2008, Unpacking the cognitive architecture of emotion processes, *Emotion* **8**(3), 341–351.

Gritti, T., Shan, C., Jeanne, V. and Braspenning, R.: 2008, Local features based facial expression recognition with face registration errors, *8th IEEE International Conference on Automatic Face Gesture Recognition, 2008. FG '08*, pp. 1–8.

Grossman, P. and Taylor, E. W.: 2007, Toward understanding respiratory sinus arrhythmia: relations to cardiac vagal tone, evolution and biobehavioral functions, *Biological psychology* **74**(2), 263–285.

Gruzelier, J. H.: 2002, A review of the impact of hypnosis, relaxation, guided imagery and individual differences on aspects of immunity and health, *Stress* **5**(2), 147–163.

Gunes, H. and Piccardi, M.: 2009, Automatic temporal segment detection and affect recognition from face and body display, *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics* **39**(1), 64–84.

Hammer, B. and Villmann, T.: 2002, Generalized relevance learning vector quantization, *Neural Networks* **15**(8-9), 1059–1068.

Hammes, J.: 1971, *De stroop kleur-woord test: handleiding*, Swets & Zeitlinger.

Healey, J. A. and Picard, R. W.: 2005, Detecting stress during real-world driving tasks using physiological sensors, *IEEE Transactions on Intelligent Transportation Systems* **6**(2), 156–166.

Heber, E., Ebert, D. D., Lehr, D., Nobis, S., Berking, M. and Riper, H.: 2013, Efficacy and cost-effectiveness of a web-based and mobile stress-management intervention for employees: design of a randomized controlled trial, *BMC Public Health* **13**(1), 1.

Hosseini, S. A., Khalilzadeh, M. A. and Changiz, S.: 2010, Emotional stress recognition system for affective computing based on bio-signals, *Journal of Biological Systems* **18**(1), 101–114.

Houtveen, J. H., Rietveld, S. and de Geus, E. J. C.: 2002, Contribution of tonic vagal modulation of heart rate, central respiratory drive, respiratory depth, and respiratory frequency to respiratory sinus arrhythmia during mental stress and physical exercise, *Psychophysiology* **39**(4), 427–436.

Jabid, T., Kabir, H. and Chae, O.: 2010a, Robust facial expression recognition based on local directional pattern, *ETRI Journal* **32**(5), 784–794.

Jabid, T., Kabir, M. H. and Chae, O.: 2010b, Facial expression recognition using local directional pattern (LDP), *2010 17th IEEE International Conference on Image Processing (ICIP)*, pp. 1605–1608.

James, W.: 1884, What is an emotion?, *Mind* **9**(34), 188–205.

Janssen, J. H., IJsselsteijn, W. A., Westerink, J. H. D. M., Tacken, P. and de Vries, G.-J.: 2013, The tell-tale heart: Perceived emotional intensity of heartbeats, *International Journal of Synthetic Emotions* **4**(1), 65–91.

Janssen, J. H., Tacken, P., de Vries, J. J. G., van den Broek, E. L., Westerink, J. H. D. M., Haselager, P. and IJsselsteijn, W. A.: 2013, Machines outperform lay persons in recognizing emotions elicited by autobiographical recollection, *Human Computer Interaction* **28**(6), 479–517.

Jones, F. and Bright, J.: 2001, *Stress: Myth, Theory and Research*, Pearson Education.

Kabir, H., Jabid, T. and Chae, O.: 2012, Local directional pattern variance (LDPv): a robust feature descriptor for facial expression recognition, *International Arab Journal of Information Technology (IAJIT)* **9**(4), 382–391.

Kanade, T., Cohn, J. F. and Tian, Y.: 2000, Comprehensive database for facial expression analysis, *Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000. Proceedings*, pp. 46–53.

Katsis, C. D., Katertsidis, N., Ganiatsas, G. and Fotiadis, D. I.: 2008, Toward emotion recognition in car-racing drivers: A biosignal processing approach, *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* **38**(3), 502–512.

Keshtkar, F. and Inkpen, D.: 2010, A corpus-based method for extracting paraphrases of emotion terms, *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, CAAGET '10, Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 35–44.

Kim, J. and André, E.: 2008, Emotion recognition based on physiological changes in music listening, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**(12), 2067–2083.

Kim, K. H., Bang, S. W. and Kim, S. R.: 2004, Emotion recognition system using short-term monitoring of physiological signals, *Medical & Biological Engineering & Computing* **42**(3), 419–427.

Kirchner, W. K.: 1958, Age differences in short-term retention of rapidly changing information, *Journal of experimental psychology* **55**(4), 352–358.

Kivimäki, M., Virtanen, M., Elovainio, M., Kouvonen, A., Väänänen, A. and Vahtera, J.: 2006, Work stress in the etiology of coronary heart disease–a meta-analysis, *Scandinavian journal of work, environment & health* **32**(6), 431–442.

Kleinginna Jr, P. R. and Kleinginna, A. M.: 1981, A categorized list of emotion definitions, with suggestions for a consensual definition, *Motivation and Emotion* **5**(4), 345–379.

Kohlish, P.: 1992, SCRGAUGE - a computer program for the detection and quantification of SCRs, *in* W. Boucsein (ed.), *Electrodermal Activity*, Plenum, New York, pp. 432–442.

Kohonen, T.: 1990, Improved versions of learning vector quantization, *International Joint Conference on Neural Networks*, Vol. 1, IEEE, pp. 545–550.

Kohonen, T.: 2001, *Self Organising Maps*, Springer, Berlin 3rd ed.

Kuechenmeister, C. A., Hain, J. D. and McClusky, H. Y.: 1970, Contingent computer averaging of evoked heart rate response to visual stimuli, *Proceedings of the 23rd annual conference on engineering in medicine and biology*, IEEE, New York, NY, USA, p. 326.

Lajevardi, S. M. and Hussain, Z. M.: 2010, Novel higher-order local autocorrelation-like feature extraction methodology for facial expression recognition, *IET Image Processing* **4**(2), 114–119.

Landis, J. R. and Koch, G. G.: 1977, The measurement of observer agreement for categorical data, *Biometrics* **33**(1), 159.

Lazarus, R.: 1991, *Emotion and Adaptation*, Oxford University Press.

Lazarus, R. S. and Folkman, S.: 1984, *Stress, Appraisal, and Coping*, Springer Publishing Company.

Lemire, D.: 2006, Streaming maximum-minimum filter using no more than three comparisons per element, *Nordic Journal of Computing* **13**(4), 328–339.

Li, Z., Imai, J. and Kaneko, M.: 2009, Facial-component-based bag of words and PHOG descriptor for facial expression recognition, *IEEE International Conference on Systems, Man and Cybernetics, 2009. SMC 2009*, pp. 1353–1358.

Lien, J. J.-J., Kanade, T., Cohn, J. F. and Li, C.-C.: 2000, Detection, tracking, and classification of action units in facial expression, *Robotics and Autonomous Systems* **31**(3), 131–146.

Lin, D.-T. and Pan, D.-C.: 2009, Integrating a mixed-feature model and multiclass support vector machine for facial expression recognition, *Integrated Computer-Aided Engineering* **16**(1), 61–74.

Lisetti, C. L. and Nasoz, F.: 2004, Using noninvasive wearable computers to recogniza human emotions from physiological signals, *Journal of Applied Signal Processing* **11**, 1672–1687.

Lisetti, C. L. and Nasoz, F.: 2005, Affective intelligent car interfaces with emotion recognition, *Proceedings of the 11th International Conference on Human Computer Interaction*, Las Vegas, USA, p. 41.

Littlewort, G., Bartlett, M. S., Fasel, I., Susskind, J. and Movellan, J.: 2006, Dynamics of facial expression extracted automatically from video, *Image and Vision Computing* **24**(6), 615–625.

Liu, N., Dellandréa, E., Chen, L., Zhu, C., Zhang, Y., Bichot, C.-E., Bres, S. and Tellez, B.: 2013, Multimodal recognition of visual concepts using histograms of textual concepts and selective weighted late fusion scheme, *Computer Vision and Image Understanding* **117**(5), 493–512.

Logan, G. and Cowan, W.: 1984, On the ability to inhibit thought and action: A theory of an act of control., *Psychological Review* **91**(3), 295–327.

Lövheim, H.: 2012, A new three-dimensional model for emotions and monoamine neurotransmitters, *Medical hypotheses* **78**(2), 341–348.

Lu, H., Wang, Z. and Liu, X.: 2006, Facial expression recognition using NKFDA method with gabor features, *The Sixth World Congress on Intelligent Control and Automation, 2006. WCICA 2006*, Vol. 2, pp. 9902–9906.

Machajdik, J. and Hanbury, A.: 2010, Affective image classification using features inspired by psychology and art theory, *Proceedings of the international conference on Multimedia*, MM '10, ACM, New York, NY, USA, pp. 83–92.

Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J. and Schwartz, P. J.: 1996, Heart rate variability standards of measurement, physiological interpretation, and clinical use, *European Heart Journal* **17**(3), 354–381.

Marsella, S., Gratch, J. and Petta, P.: 2010, Computational models of emotion, *in* K. R. Scherer, T. Bänziger and E. B. Roesch (eds), *Blueprint for affective computing: A sourcebook*, Series in Affective Science, Oxford University Press.

Martens, D., Baesens, B., van Gestel, T. and Vanthienen, J.: 2007, Comprehensible credit scoring models using rule extraction from support vector machines, *European Journal of Operational Research* **183**(3), 1466–1476.

Mathewson, K. J., Jetha, M. K., Drmic, I. E., Bryson, S. E., Goldberg, J. O., Hall, G. B., Santesso, D. L., Segalowitz, S. J. and Schmidt, L. A.: 2010, Autonomic predictors of stroop performance in young and middle-aged adults, *International Journal of Psychophysiology* **76**(3), 123–129.

McEwen, B. S., Goodman, H. M. and American Physiological Society (1887-): 2001, *Coping with the environment: neural and endocringe mechanisms*, number IV in *Handbook of Physiology (Section 7: The endocrine system)*, Oxford University Press, New York.

McNames, J. and Aboy, M.: 2006, Reliability and accuracy of heart rate variability metrics versus ECG segment duration, *Medical and Biological Engineering and Computing* **44**(9), 747–756.

McRae, K., Misra, S., Prasad, A. K., Pereira, S. C. and Gross, J. J.: 2012, Bottom-up and top-down emotion generation: implications for emotion regulation, *Social cognitive and affective neuroscience* **7**(3), 253–262.

Mehrabian, A. and Russell, J. A.: 1974, *An approach to environmental psychology*, M.I.T. Press.

Meir, R.: 1995, Empirical risk minimization versus maximum-likelihood estimation: a case study, *Neural computation* **7**, 144–157.

Meuleman, B. and Scherer, K.: 2013, Nonlinear appraisal modeling: An application of machine learning to the study of emotion production, *IEEE Transactions on Affective Computing* **4**(4), 398–411.

Neural Networks Research Centre, Helsinki: 2002, Bibliography on the self-organizing maps (SOM) and learning vector quantization (LVQ), *Otaniemi: Helsinki Univ. of Technology. Available on-line:* `http://liinwww.ira.uka.de/bibliography/Neural/SOM.LVQ.html` .

Nichols, A. L. and Maner, J. K.: 2008, The good-subject effect: investigating participant demand characteristics, *The Journal of General Psychology* **135**(2), 151–165.

Nikitidis, S., Tefas, A., Nikolaidis, N. and Pitas, I.: 2011, Facial expression recognition using clustering discriminant non-negative matrix factorization, *2011 18th IEEE International Conference on Image Processing (ICIP)*, pp. 3001–3004.

Nummenmaa, L., Glerean, E., Hari, R. and Hietanen, J. K.: 2013, Bodily maps of emotions, *Proceedings of the National Academy of Sciences* .

Ochsner, K. N., Ray, R. R., Hughes, B., McRae, K., Cooper, J. C., Weber, J., Gabrieli, J. D. E. and Gross, J. J.: 2009, Bottom-up and top-down processes in emotion generation: common and distinct neural mechanisms, *Psychological science* **20**(11), 1322–1331.

Ojala, T., Pietikäinen, M. and Mäenpää, T.: 2002, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(7), 971–987.

Orrite, C., Gañán, A. and Rogez, G.: 2009, HOG-Based decision tree for facial expression classification, *in* D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, H. Araujo, A. M. Mendonça, A. J. Pinho and M. I. Torres (eds), *Pattern Recognition and Image Analysis*, Vol. 5524, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 176–183.

Ortony, A., Clore, G. L. and Collins, A.: 1988, *The Cognitive Structure of Emotions*, Cambridge University Press.

Overbeek, T. J. M., Van Boxtel, A. and Westerink, J. H. D. M.: 2007, Development of an Emotion-Eliciting stimulus set: Results of emotional pictures and film fragments ratings, *Technical Report PR-TN 2007/00574*, Philips Research Technical Note PR-TN 2007/00574.

Pantic, M. and Patras, I.: 2006, Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* **36**(2), 433–449.

Pantic, M., Sebe, N., Cohn, J. F. and Huang, T.: 2005, Affective multimodal human-computer interaction, *Proceedings of the 13th Annual ACM International Conference on Multimedia*, MULTIMEDIA '05, ACM, New York, NY, USA, pp. 669–676.

Penttilä, J., Helminen, A., Jartti, T., Kuusela, T., Huikuri, H. V., Tulppo, M. P., Coffeng, R. and Scheinin, H.: 2001, Time domain, geometrical and frequency domain analysis of cardiac vagal outflow: effects of various respiratory patterns, *Clinical Physiology* **21**(3), 365–376.

Peter, C. and Beale, R.: 2008, *Affect and Emotion in Human-Computer Interaction - From Theory to Applications*, number 4868 in *Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg.

Picard, R. W.: 1995, Affective Computing, *Technical report*, MIT.

Picard, R. W. and Scheirer, J.: 1999, The galvactivator: A glove that senses and communicates skin conductivity, *Proceedings from the 9th International Conference on Human-Computer Interaction*, New Orleans, LA, pp. 1538–1542.

Picard, R. W., Vyzas, E. and Healey, J.: 2001, Toward machine emotional intelligence: Analysis of affective physiological state, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**(10), 1175–1191.

Poli, S., Sarlo, M., Bortoletto, M., Buodo, G. and Palomba, D.: 2007, Stimulus-Preceding negativity and heart rate changes in anticipation of affective pictures, *International Journal of Psychophysiology* **65**(1), 32–39.

Poor, H. V.: 1994, *An Introduction to Signal Detection and Estimation*, Springer.

Poursaberi, A., Noubari, H., Gavrilova, M. and Yanushkevich, S. N.: 2012, Gauss-Laguerre wavelet textural feature fusion with geometrical information for facial expression identification, *EURASIP Journal on Image and Video Processing* **2012**(1), 17.

Ptaszynski, M., Dybala, P., Shi, W., Rzepka, R. and Araki, K.: 2009, Towards context aware emotional intelligence in machines: computing contextual appropriateness of affective states, *Proceedings of the 21st international joint conference on Artificial intelligence*, IJCAI'09, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 1469–1474.

Rani, P., Liu, C., Sarkar, N. and Vanman, E.: 2006, An empirical study of machine learning techniques for affect recognition in human-robot interaction, *Pattern Analysis & Applications* **9**(1), 58–69.

Reents, G. and Urbanczik, R.: 1998, Self averaging and on-line learning, *Phys. Rev. Letter* **80**, 5445–5448.

Rifkin, R. and Klautau, A.: 2004, In defense of one-vs-all classification, *The Journal of Machine Learning Research* **5**, 101–141.

Rosch, P. J.: 2001, The quandary of job stress compensation, *Health & Stress* pp. 1–4.

Russell, J. A.: 2003, Core affect and the psychological construction of emotion, *Psychological Review* **110**, 145–172.

Russell, J. A. and Mehrabian, A.: 1977, Evidence for a three-factor theory of emotions, *Journal of Research in Personality* **11**(3), 273–294.

Saad, D. (ed.): 1999, *Online learning in neural networks*, Cambridge University Press, Cambridge, UK.

Saad, D. and Rattray, M.: 1997, Globally optimal parameters for on-line learning in multilayer neural networks, *Phys. Rev. Lett.* **79**, 2578–2581.

Saad, D. and Solla, S. A.: 1995, On-line learning in soft committee machines, *Phys. Rev. E* **52**, 4225–4243.

Saha, A. and Wu, Q. M. J.: 2010, Facial expression recognition using curvelet based local binary patterns, *2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp. 2470–2473.

Sanchez, A., Ruiz, J. V., Moreno, A. B., Montemayor, A. S., Hernandez, J. and Pantrigo, J. J.: 2010, Differential optical flow applied to automatic facial expression recognition, *Neurocomputing* **74**(8), 1272–1282.

Sato, A. and Yamada, K.: 1995, Generalized learning vector quantization, *NIPS* pp. 423–429.

Scherer, K. R.: 1993, Studying the Emotion-Antecedent appraisal process: An expert system approach, *Cognition and Emotion* **7**(3/4), 325–355.

Scherer, K. R.: 2001, Appraisal considered as a process of multilevel sequential checking, *Appraisal processes in emotion: Theory, Methods, Research* pp. 92–120.

Scherer, K. R.: 2005, What are emotions? and how can they be measured?, *Social Science Information* **44**(4), 695–729.

Scherer, K. R., Dan, E. S. and Flykt, A.: 2006, What determines a feeling's position in affective space? a case for appraisal, *Cognition and Emotion* **20**(1), 92–113.

Scherer, K. R., Schorr, A. and Johnstone, T. (eds): 2001, *Appraisal processes in emotion: Theory, methods, research*, Oxford University Press, New York.

Schneider, P., Biehl, M. and Hammer, B.: 2009a, Adaptive relevance matrices in learning vector quantization, *Neural Computation* **21**(12), 3532–3561.

Schneider, P., Biehl, M. and Hammer, B.: 2009b, Distance learning in discriminative vector quantization, *Neural computation* **21**(10), 2942–2969.

Sebe, N.: 2005, Mona lisa 'happy', computer finds, `http://news.bbc.co.uk/2/hi/entertainment/4530650.stm`.

Sebe, N.: 2006, Mona lisa: Smiling? – computer scientists develop software that evaluates facial expressions, `http://www.sciencedaily.com/videos/2006/0811-mona_lisa_smiling.htm`.

Seo, S. and Obermayer, K.: 2003, Soft learning vector quantization, *Neural Computation* **15**, 1589–1604.

Seo, S. and Obermayer, K.: 2006, Dynamic hyper parameter scaling method for lvq algorithms, *International Joint Conference on Neural Networks, Vancouver, Canada*.

Shan, C., Gong, S. and McOwan, P. W.: 2009, Facial expression recognition based on local binary patterns: A comprehensive study, *Image and Vision Computing* **27**(6), 803–816.

Sharma, N. and Gedeon, T.: 2012, Objective measures, sensors and computational techniques for stress recognition and classification: A survey, *Computer Methods and Programs in Biomedicine* **108**(3), 1287–1301.

Shaver, P., Schwartz, J., Kirson, D. and O'Connor, C.: 2001, Emotional knowledge: Further exploration of a prototype approach, *in* G. Parrott (ed.), *Emotions in Social Psychology: Essential Readings*, Psychology Press, Philadelphia, PA, pp. 26–56.

Sinha, R. and Parsons, O. A.: 1996, Multivariate response patterning of fear, *Cognition and Emotion* **10**(2), 173–198.

Sitskoorn, M. M., van Boxtel, G. J. M., Geurdes, J. I. M., Vernon, D. J., Denissen, A. J. M., Holten, V. and Jaeger, M.: 2009, Effectiveness study on a lean-backward audio based neurofeedback training (nft) to enhance cognitive performance, mood and reduce stress of healthy subjects. Unpublished study protocol.

Smith, C. A.: 1989, Dimensions of appraisal and physiological response in emotion., *Journal of Personality and Social Psychology* **56**(3), 339–353.

Smith, C. A. and Lazarus, R. S.: 1990, Emotion and adaptation, *in* L. A. Pervin (ed.), *Handbook of personality: Theory and research*, Guilford, New York.

Sobol-Shikler, T. and Robinson, P.: 2010, Classification of complex information: Inference of co-occurring affective states from their expressions in speech, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32**(7), 1284–1297.

Song, M., Tao, D., Liu, Z., Li, X. and Zhou, M.: 2010, Image ratio features for facial expression recognition application, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* **40**(3), 779–788.

Stern, R. M., Ray, W. J. and Quigley, K. S.: 2001, *Psychophysiological Recording*, Oxford University Press.

Strickert, M., Hammer, B., Villmann, T. and Biehl, M.: 2013, Regularization and improved interpretation of linear data mappings and adaptive distance measures, *Proc. IEEE SSCI 2013*, Singapore, pp. 10–17.

Swerts, M. and Krahmer, E.: 2008, Facial expression and prosodic prominence: Effects of modality and facial area, *Journal of Phonetics* **36**(2), 219–238.

Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology: 1996, Heart rate variability standards of measurement, physiological interpretation, and clinical use, *Circulation* **93**(5), 1043–1065.

Thomson Reuters: 2014, Web of science, `http://apps.webofknowledge.com/`.

Tian, Y.-L.: 2004, Evaluation of face resolution for expression analysis, *Conference on Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04*, pp. 82–82.

Tickle, A. B., Andrews, R., Golea, M. and Diederich, J.: 1998, The truth will come to light: directions and challenges in extracting the knowledge embedded within trained artificial neural networks, *IEEE Transactions on Neural Networks* **9**(6), 1057–1068.

Valdez, P. and Mehrabian, A.: 1994, Effects of color on emotions, *Journal of Experimental Psychology: General* **123**(4), 394–409.

van Boxtel, G. J. M., Denissen, A. J. M., Jäger, M., Vernon, D., Dekker, M. K. J., Mihajlović, V. and Sitskoorn, M. M.: 2012, A novel self-guided approach to alpha activity training, *International Journal of Psychophysiology* **83**(3), 282–294.

van den Broek, E. L., Janssen, J. H., Westerink, J. H. D. M. and Healey, J. A.: 2009, Prerequisites for affective signal processing (asp), *in* P. E. ao and A. Veloso (eds), *Biosignals 2009: Proceedings of the Second International Conference on Bio-Inspired Systems and Signal Processing*, INSTICC Press, Portugal, pp. 426–433.

van den Broek, E. L., Lisý, V., Janssen, J. H., Westerink, J. H. D. M., Schut, M. H. and Tuinenbreijer, K.: 2010, Affective man-machine interface: Unveiling human emotions through biosignals, *in* A. Fred, J. Filipe and H. Gamboa (eds), *Biomedical Engineering Systems and Technologies: BIOSTEC2009 Selected Revised papers*, Vol. 52 of *Communications in Computer and Information Science*, Springer, Berlin/Heidelberg, Germany, pp. 21–47.

van den Broek, E. L., van der Sluis, F. and Dijkstra, T.: 2011, Telling the story and re-living the past: How speech analysis can reveal emotions in post-traumatic stress disorder (ptsd) patients, *in* J. H. D. M. Westerink, M. Krans and

M. Ouwerkerk (eds), *Sensing Emotions: The impact of context on experience measurements*, Vol. 12 of *Philips Research Book Series*, Dordrecht, The Netherlands: Springer Science + Business Media B.V., pp. 153–180.

van den Broek, E. L., van der Zwaag, M. D., Healey, J. A., Janssen, J. H. and Westerink, J. H. D. M.: 2010, Prerequisites for affective signal processing (asp) - part iv, *in* J. Kim and P. Karjalainen (eds), *Proceedings of the 1st International Workshop on Bio-inspired Human-Machine Interfaces and Healthcare Applications - B-Interface 2010*, INSTICC Press, Portugal, pp. 59–66.

van der Zwaag, M. D., Janssen, J. H. and Westerink, J. H. D. M.: 2012, Directing physiology and mood through music: Validation of an affective music player, *IEEE Transactions on Affective Computing* **4**(1), 57–68.

van Dooren, M., de Vries, J. J. G. and Janssen, J. H.: 2012, Emotional sweating across the body: Comparing 16 different skin conductance measurement locations, *Physiology & Behavior* **106**(2), 298–304.

van Kuilenburg, H., den Uyl, M. J., Israel, M. L. and Ivan, P.: 2008, Advances in face and gesture analysis, *Measuring Behavior 2008* pp. 371–372.

van Reekum, C., Johnstone, T., Banse, R., Etter, A., Wehrle, T. and Scherer, K.: 2004, Psychophysiological responses to appraisal dimensions in a computer game, *Cognition & Emotion* **18**(5), 663–688.

Vapnik, V.: 1998, *Statistical learning theory*, Wiley.

Veloso, L. R., Carvalho, J. M., Cavalvanti, C. S. V. C., Moura, E. S., Coutinho, F. L. and Gomes, H. M.: 2007, Neural network classification of photogenic facial expressions based on fiducial points and gabor features, *in* D. Mery and L. Rueda (eds), *Advances in Image and Video Technology*, Vol. 4872, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 166–179.

Vuksanović, V. and Gal, V.: 2007, Heart rate variability in mental stress aloud, *Medical Engineering & Physics* **29**(3), 344–349.

Wan, S. and Aggarwal, J. K.: 2014, Spontaneous facial expression recognition: A robust metric learning approach, *Pattern Recognition* **47**(5), 1859–1868.

Wang, J. and Yin, L.: 2007, Static topographic modeling for facial expression recognition and analysis, *Computer Vision and Image Understanding* **108**(1-2), 19–34.

Wang, Z. and Ruan, Q.: 2010, Facial expression recognition based orthogonal local fisher discriminant analysis, *2010 IEEE 10th International Conference on Signal Processing (ICSP)*, pp. 1358–1361.

Watson, D. and Tellegen, A.: 1985, Toward a consensual structure of mood, *Psychological Bulletin* **98**(2), 219–235.

Westerink, J., Ouwerkerk, M., de Vries, G.-J., de Waele, S., van den Eerenbeemd, J. and van Boven, M.: 2009, Emotion measurement platform for daily life situations, *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pp. 1–6.

Westerink, J., van Beek, W., Daemen, E., Janssen, J., de Vries, G.-J. and Ouwerkerk, M.: 2014, The vitality bracelet: Bringing balance to your life with psychophysiological measurements, *in* S. H. Fairclough and K. Gilleade (eds), *Advances in Physiological Computing*, Springer London, London, pp. 197–209.

Westerink, J., van Boxtel, A., IJsselsteijn, W., Janssen, J., Ouwerkerk, M., Overbeek, T., de Vries, G.-J., Slovak, P., van der Zwaag, M. and Fitzpatrick, G.: 2012, Unobtrusive emotion sensing in everyday life, *in* A. J. Spink, F. Grieco, O. E. Krips, L. W. S. Looijens, L. P. J. J. Noldus and P. H. Zimmerman (eds), *Proceedings of Measuring Behavior 2012, 8th International Conference on Methods and Techniques in Behavioral Research*, Utrecht, the Netherlands, p. 332.

Wijsman, J., Grundlehner, B., Liu, H., Hermens, H. and Penders, J.: 2011, Towards mental stress detection using wearable physiological sensors, *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society,EMBC*, pp. 1798–1801.

Witoelar, A., Biehl, M., Ghosh, A. and Hammer, B.: 2008, Learning dynamics and robustness of vector quantization and neural gas, *Neurocomputing* **71**, 1210–1219.

Witoelar, A. W., Ghosh, A., de Vries, J. J. G., Hammer, B. and Biehl, M.: 2011, Window-based example selection in learning vector quantization, *Neural Computation* **22**(11), 2924–2961.

Wright, C. E., O'Donnell, K., Brydon, L., Wardle, J. and Steptoe, A.: 2007, Family history of cardiovascular disease is associated with cardiovascular responses to stress in healthy young men and women, *International Journal of Psychophysiology* **63**(3), 275–282.

Wu, S., Falk, T. H. and Chan, W.-Y.: 2011, Automatic speech emotion recognition using modulation spectral features, *Speech Communication* **53**(5), 768–785.

Xiao, R., Zhao, Q., Zhang, D. and Shi, P.: 2011, Facial expression recognition on multiple manifolds, *Pattern Recognition* **44**(1), 107–116.

Xu, Q., Zhang, P., Pei, W., Yang, L. and He, Z.: 2006, A facial expression recognition approach based on confusion-crossed support vector machine tree, *International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2006. IIH-MSP '06*, pp. 309–312.

Xu, Q., Zhang, P., Pei, W., Yang, L. and He, Z.: 2007, An automatic facial expression recognition approach based on confusion-crossed support vector machine tree, *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007*, Vol. 1, pp. I–625–I–628.

Xu, Q., Zhang, P., Yang, L., Pei, W. and He, Z.: 2007, A facial expression recognition approach based on novel support vector machine tree, *in* D. Liu, S. Fei, Z. Hou, H. Zhang and C. Sun (eds), *Advances in Neural Networks - ISNN 2007*, Vol. 4493, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 374–381.

Xu, T., Zhou, J. and Wang, Y.: 2011, A variation of local directional pattern and its application for facial expression recognition, *in* T.-H. Kim, H. Adeli, C. Ramos and B.-H. Kang (eds), *Signal Processing, Image Processing and Pattern Recognition*, Vol. 260, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 36–47.

Xue, G. and Youwei, Z.: 2006, Facial expression recognition based on the difference of statistical features, *2006 8th International Conference on Signal Processing*, Vol. 3.

Yacoob, Y. and Davis, L. S.: 2006, Recognizing human facial expressions from long image sequences using optical flow, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18**(6), 636–642.

Young, P. T.: 1973, Feeling and emotion, *in* B. B. Wolman (ed.), *Handbook of general psychology*, Prentice-Hall, Englewood Cliffs, New Jersey.

Zafeiriou, S. and Pitas, I.: 2008, Discriminant graph structures for facial expression recognition, *IEEE Transactions on Multimedia* **10**(8), 1528–1540.

Zavaschi, T. H. H., Britto Jr., A. S., Oliveira, L. E. S. and Koerich, A. L.: 2013, Fusion of feature sets and classifiers for facial expression recognition, *Expert Systems with Applications* **40**(2), 646–655.

Zhai, J. and Barreto, A.: 2006, Stress detection in computer users based on digital signal processing of noninvasive physiological variables, *Annual International Conference of the IEEE Engineering in Medicine and Biology Society.*, Vol. 1, pp. 1355–1358.

Zhai, J., Barreto, A. B., Chin, C. and Li, C.: 2005, Realization of stress detection using psychophysiological signals for improvement of human-computer interactions, *IEEE SoutheastCon, 2005. Proceedings*, pp. 415–420.

Zhang, L. and Tjondronegoro, D.: 2009, Selecting, optimizing and fusing 'salient' gabor features for facial expression recognition, *in* D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, C. S. Leung, M. Lee and J. H. Chan (eds), *Neural Information Processing*, Vol. 5863, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 724–732.

Zhang, L. and Tjondronegoro, D.: 2010, Improving the performance of facial expression recognition using dynamic, subtle and regional features, *in* D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, C. S. Leung, M. Lee and J. H. Chan (eds), *Neural Information Processing. Models and Applications*, Vol. 6444, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 582–589.

Zhao, X. and Zhang, S.: 2011, Facial expression recognition based on local binary patterns and kernel discriminant isomap, *Sensors (Basel, Switzerland)* **11**(10), 9573–9588.

Zhao, X. and Zhang, S.: 2012, Facial expression recognition using local binary patterns and discriminant kernel locally linear embedding, *EURASIP Journal on Advances in Signal Processing* **2012**(1), 20.

Zhi, R., Flierl, M., Ruan, Q. and Kleijn, W. B.: 2011, Graph-preserving sparse nonnegative matrix factorization with application to facial expression recognition, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* **41**(1), 38–52.

Zhi, R. and Ruan, Q.: 2008, A comparative study on region-based moments for facial expression recognition, *Congress on Image and Signal Processing, 2008. CISP '08*, Vol. 2, pp. 600–604.

Zhi, R., Ruan, Q. and Miao, Z.: 2008, Fuzzy discriminant projections for facial expression recognition, *19th International Conference on Pattern Recognition, 2008. ICPR 2008*, pp. 1–4.

Zhou, J., Xu, T., Wang, Y., Gao, L. and Yang, R.: 2011, A novel feature extraction for facial expression recognition via combining the curvelet and LDP, *in* J. Kacprzyk and R. Lee (eds), *Computer and Information Science 2011*, Vol. 364, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 35–46.

Zilu, Y., Jingwen, L. and Youwei, Z.: 2006, Facial expression recognition based on classifier combinations, *2006 8th International Conference on Signal Processing*, Vol. 3.

Zysset, S., Müller, K., Lohmann, G. and von Cramon, D. Y.: 2001, Color-word matching stroop task: Separating interference and response conflict, *NeuroImage* **13**(1), 29–36.