Technische Universität München

Lehrstuhl für Steuerungs- und Regelungstechnik
Univ.-Prof. Dr.-Ing./Univ. Tokio Martin Buss

Pattern Recognition Algorithms for Gait Analysis with Application to Affective Computing

Michelle E. Karg

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

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Gordon Cheng, Ph. D.

Prüfer der Dissertation:

1. TUM Junior Fellow Dr.-Ing. Kolja Kühnlenz

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

Die Dissertation wurde am 04.10.2011 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 11.01.2012 angenommen.

Foreword

This thesis summarizes my work as research assistant at the Institute of Automatic Control Engineering (LSR), TU München, from 2007 to 2011.

First of all, I would like to express my profound gratitude towards my doctoral advisor Dr.-Ing. Kolja Kühnlenz who always supported me with his guidance, trust, and experience. I could always count on his support and he gave me also the liberty to work out own ideas. I sincerely thank Univ.-Prof. Dr.-Ing./Univ. Tokio Martin Buss for always supporting me with his confidence and for giving me the opportunity to start on a fascinating topic and in an inspiring working environment.

Furthermore, I would like to thank my collaborators Dr.rer.nat. Wolfgang Seiberl, Dr. phil. Ferdinand Tusker, and Prof. Dr. Ansgar Schwirtz in the CoTeSys project 151, Dipl.-Psych. Conny Wendt and Prof. Dr. Berthold Färber in the CoTeSys project 126, Dr. Stephan Haug for support with statistics, and Dipl.-Psych. Eva Wiese and Dr. Jan Zwickel during the eye tracking study.

I would also like to thank all the students who contributed to this project: S. Abdurahman, B. Buchholz, L. Cao, R. Jenke, B. Karg, S. Önder, T. Schindl, T. Schupp, M. Schwimmbeck, H. Soufi, C. Spies, S. Toprak, and X. Wang. They did a very good job in their internships, seminars, student jobs, bachelor, or master theses and I really enjoyed supervising and working with them.

To all my collegues and the staff at the LSR, I am deeply grateful for working together with you, for your help whenever I needed it, for taking part at the numerous experiments, and, of course, for waiting until I ate my last piece in the Mensa. Special thanks goes to my friends and collegues Andrea Bauer, Barbara Gonsior, Carolina & Benjamin Passenberg, Harald Voit, and Tingting Xu for answering all the questions I had at the beginning, for all our discussions, for carefully proof-reading my thesis, and for the great time we spent together.

I greatly appreciate the technical support from Mr. Jaschik, Mr. Stoeber, and our admins, and the help in administrative issues from Mrs. Schmid, Mrs. Werner, and Mrs. Renner.

Finally, I would like to thank my parents, my grandparents, my sister and Fam. Scharfenberger for always supporting me, for everything they did for me, and for giving me the possibility to be where I am now. Most of all, I would like to thank Christian Scharfenberger for always standing at my side, for encouraging me whenever I needed it, and for sharing all the impressions of being a doctoral student with me.

München, September 2011 Michelle Karg

To The Reader

...

Abstract

Affective computing increasingly gains interest in human-robot interaction (HRI). Within this research area, a large part of the studies concentrates on facial expressions and variations in speech as modalities. These modalities are especially suited during an interaction, but they may not be sufficient in other situations such as recognizing the atmosphere at distance and adapting the robot to the interaction partner beforehand. For this purpose, this thesis aims to investigate the capability of the daily human motion gait for automatically recognizing and expressing emotions in HRI.

At first glance, it appears that walking follows a simple principle. Nevertheless, mathematical modeling indicates that a complex mechanism underlies human walking and, hence, a detailed analysis of the human gait is challenging. The gait is highly individual and affected by many factors such as physique, age, gender, and emotions. Furthermore, recorded gait databases are characterized by highly dimensional, temporally dependent, highly variable, and nonlinear data vectors. Due to these reasons, automatic recognition of affect is a challenge for pattern recognition algorithms.

Reviewing psychological studies leads to the conclusion that humans express emotions automatically in the way they walk. Considering the applicability to affective computing, several further experiments are conducted on human perception of emotions. Results of these experiments evaluate the suitability of the investigated gait databases, serve as reference for the performance of pattern recognition algorithms, and give new insights on the human perception mechanism.

Traditionally, inferential statistics is applied for gait analysis in medicine and biomechanics. Within this work, predictive and inferential statistics are theoretically and numerically compared. Similarities as well as dissimilarities are elaborated. Mathematical relations between these methods are derived, in particular, estimating classification rates from reported test statistics.

Different feature extraction techniques and static as well as dynamic classification methods are compared for the recognition of emotions in gait patterns. For a small number of training samples and highly dimensional feature vectors, it is derived that the decision borders of a nearest neighbor classifier coincide with the decision borders of a support vector machine (SVM) if linear discriminant analysis (LDA) is used for dimension reduction. Besides developing well performing algorithms, crucial issues in automatic emotion recognition are the definition of the term emotion, person dependency of the recognition performance, and the ground truth. For this purpose, two emotion models are compared regarding their suitability to study gait as modality in affective computing. Furthermore, algorithms trained for individuals are compared with algorithms not considering the identity of the walker. This work uses therefor databases which are based on acted or elicited emotions. Finally, the achieved recognition rates are compared with gender and identity recognition.

For robots and virtual characters, affective computing does not only consider automatic emotion recognition but also expression of emotions; therefore, expressive gait patterns for robots are developed and evaluated for a hexapod.

In summary, this work which deals about the suitability of gait for affective computing comes to the conclusion that even though gait is a highly individual motion pattern, emotions are recognized above chance level. This work contributes to the state of the art

by exploring various facets of pattern recognition algorithms for gait analysis and studying gait as modality for affective computing. It provides valuable insights concerning this topic and opens various perspectives for future work.

Zusammenfassung

Affective Computing gewinnt zunehmend an Bedeutung in der Mensch-Roboter Interaktion (HRI), weil diese menschliche Emotionen berücksichtigt. Ein Großteil der Studien auf diesem Forschungsgebiet konzentriert sich auf die Mimik und Variationen in der Sprache als Modalität für emotionale HRI. Diese Modalitäten sind insbesondere während der Interaktion geeignet. Jedoch treten Situationen auf, in welchen diese Modalitäten unzureichend sind, z.B. um die Stimmung des Interaktionspartners von Weitem zu erkennen, sodass der Roboter die Interaktion dementsprechend eröffnet. Deswegen untersucht diese Arbeit die Eignung des menschlichen Gangs zur Erkennung und Expression von Emotionen in HRI.

Auf den ersten Blick scheint der menschliche Gang einem einfachen Prinzip zu folgen. Eine detailreichere, mathematische Modellierung zeigt jedoch auf, dass dem menschlichen Gang ein komplexer Mechanismus zu Grunde liegt. Somit stellt eine detailierte Analyse des menschlichen Gangs eine Herausforderung für technische Applikationen dar. Der persönliche Gang ist individuell und wird durch viele Faktoren wie Körperbau, Alter, Geschlecht, und auch die emotionale Stimmung beeinflusst. Weiterhin sind aufgenommene Gangdaten dadurch charakterisiert, dass die Datenvektoren eine hohe Dimension haben, sowohl zeitabhängig als auch sehr variabel sind, und Nichtlinearitäten beinhalten.

Eine Anzahl an psychologischen Studien belegt, dass Menschen im Gang Emotionen ausdrücken und diese erkennbar sind. Im Hinblick auf die mögliche Anwendung von Emotionen in HRI, werden verschiedene weitere Experimente durchgeführt, welche Aufschluss über die menschliche Perzeption von Emotionen geben. Für einen späteren Vergleich mit Verfahren der automatischen Mustererkennung, wird dabei die menschliche Erkennungsrate bestimmt. Weiterhin wird die Eignung der verwendeten Datensätze überprüft und verschiedene Mechanismen bei der Wahrnehmung von Emotionen basierend auf der Beobachtung des Gangs untersucht.

In der Medizin und Biomechanik wird die statistische Inferenz herangezogen um Gangdaten zu analysieren. Innerhalb des letzten Jahrzehntes fanden auch verschiedene Verfahren der automatischen Mustererkennung in der Ganganalyse Anwendung. Im Rahmen dieser Arbeit wird jeweils eine Auswahl an Methoden der statistischen Inferenz und der automatischen Mustererkennung miteinander verglichen, um Ähnlichkeiten, Gemeinsamkeiten und Differenzen herauszuarbeiten. Des Weiteren werden mathematische Zusammenhänge hergeleitet um Klassifkationsraten von dokumentierten Testgrößen abzuschätzen.

Basierend auf der Beobachtung des Gangs, werden verschiedene Dimensionsreduktionsverfahren und Klassifikationsmethoden auf deren Eignung untersucht Rückschlüsse auf den emotionalen Zustand zu erlauben. In diesem Rahmen wird mathematisch hergeleitet, dass das Nearest Neighbor Verfahren die gleichen Entscheidungsgrenzen berechnet wie eine Support Vector Machine (SVM), wenn das Dimensionsreduktionsverfahren Linear Discriminant Analysis (LDA) im Falle geringer Trainingsdaten und hochdimensionaler Merkmalsräume angewandt wird. Neben der Entwicklung von Mustererkennungsalgorithmen, sind weitere kritische Punkte bei der Erkennung von Emotionen zu berücksichtigen: die Personenabhängigkeit, die Definition des Begriffs Emotion und der Wahrheitsgehalt. Zu diesem Zwecke werden zwei Emotionsmodelle auf deren Eignung, Emotionen am Gang zu erkennen, untersucht. Desweitern werden für einzelne Personen trainierte Algorithmen mit Algorithmen verglichen, welche personenunabhängig klassifizieren. Diese Arbeit verwendet

hierbei Datensätze, die auf gespielten oder induzierten Emotionen basieren. Abschließend werden die erreichten Erkennungsraten mit der Geschlechts- und Identitätserkennung verglichen.

Nachdem die Betrachtung von Emotion in HRI nicht nur die Erkennung von menschlichen Emotionen beinhaltet, sondern auch dass Roboter emotional interagieren, werden abschließend emotionale Gangmuster für einen sechsbeinigen Roboter entwickelt und evaluiert.

Zusammenfassend kommt diese Arbeit, welche die Eignung des Gangs zur Erkennung und Expression von Emotionen in HRI analysiert, zu dem Schluss dass die Erkennung von Emotionen am Gang über zufällige Zuordnung hinausgeht. Jedoch wird der Ausdruck stark durch den persönlichen Gangstil beeinflusst. Diese Arbeit trägt zu dem bisherigen Stand der Technik mit der Erforschung von verschiedenen Facetten der automatischen Mustererkennung für die Ganganalyse und der Untersuchung des Gangs als Modalität für emotionale HRI bei. Sie bietet wertvolle Einsichten bezüglich dieses Themas und öffnet weitere Perspektiven für zukünftige Forschung.

Contents

Notations

Abbreviations

ANOVA	Analysis of Variance
ANN	Artificial Neural Networks
CMC	Cumulative Match Characteristic
CDHMM	Continuous Density Hidden Markov Modell
DDHMM	Discrete Density Hidden Markov Modell
CMC	Cumulative Match Characteristic
DOF	Degree of Freedom
FACS	Facial Action Coding System
FED	Frame to Exemplar Distance
FT	Fourier Transformation
GDA	General Discriminant Analysis
GMM	Gaussian Mixture Model
HCI	Human Computer Interaction
HMM	Hidden Markov Model
HoV	Homogeneity of Variance
HRI	Human Robot Interaction
KNN	k-Nearest Neighbor
KPCA	Kernel Principal Component Analysis
LDA	Linear Discriminant Analysis
MANOVA	Multivariate Analysis of Variance
NHST	Null hypothesis significance testing
NN	Nearest Neighbor
PAD	Pleasure-Arousal-Dominance
PC	Principal Component
PCA	Principal Component Analysis
PLW	Point-Light-Walker
PWM	Pulse-Width Modulation
QZ	Generalized Schur decomposition
SAM	Self-Assessment Manikin
SBS	Sequential Bachkward Search
SFS	Sequential Forward Search
SVM	Support Vector Machine
VSC	Velocity, Stride Length and Cadence

Conventions

Scalars, Vectors, Matrices, and Functions

x	Scalar
$\mathbf{x}$	Vector
$\mathbf{X}$	Matrix
$\lvert \cdot \rvert$	Absolute Value
$\lVert \cdot \rVert$	Norm
$\mathbb{R}$	Set
$f(\cdot)$	Scalar function
$\boldsymbol{f}(\cdot)$	Vector function
$\exp(\cdot)$	Exponential function
$diag(\cdot)$	Diagonal matrix
$\log(\cdot)$	Natural logarithm
$\sin(\cdot)$	Sinus function
$\#$	Number of
Δ	Difference

Subscripts and Superscripts

$(\cdot)^{-1}$	Inverse
$(\cdot)^{T}$	Transpose
$\hat{x}$	Estimated value
$\overline{x}$	Mean value

Symbols

Perception of Emotions from Gait Patterns

A	Amplitude
a	Angle
C	Chance level
df	Degree of freedom
F	F-distributed variable
p	p-value
$\bar{R}$	Average recognition rate
f	Frequency
t	Time
χ^2	χ^2-distributed variable
η^2	Effect size
Φ	Phase
ϕ_i	i-th joint angle

σ_W	Within walkers variance
σ_P	Within participants variance

Comparison of Inferential and Predictive Statistics in the Context of Gait Analysis

$\mathcal{C}_i$	State i of Nature
c	Number of classes
D	Difference of paired samples
D_{max}	Normalized maximum range
d	Cohen's index of a data set
F	F-distributed variable
F_{ν_d,ν_n}	Fisher-distribution with degree of freedom of the denominator ν_d and the nominator ν_n
f	f index
f_p	f index of the population
$f(\cdot)$	decision function
$g_i(\cdot)$	Decision boarder
H_0	Null hypothesis
H_A	Alternative hypothesis
h	Parameter
$l(\cdot,\cdot)$	Likelihood
MS_{BG}	Mean of between-class sum of square
MS_T	Mean of total sum of squares
MS_{WG}	Mean of within-class sum of squares
$\mathcal{N}(\cdot,\cdot)$	Normal distribution
N	Number of samples of the complete data set
n_i	Number of samples belonging to class i
$P_E^\star$	Bayes error rate
p	p-value
$\boldsymbol{p}$	Class representative
$P(\cdot)$	Probability
$P(\cdot\|\cdot)$	Conditional probability
$p(\cdot\|\cdot)$	Conditional probability density function
Q	Objective function
R	Classification rate, recognition rate, accuracy
$\bar{R}$	Average classification rate, average recognition rate, average accuracy
r	Effect size correlation
s	Standard deviation of a data set
s^2	Variance of a data set
s_e^2, s_p	Pooled variance estimate
SS_{BG}	Between-class sum of squares
SS_T	Total sum-of-squares
SS_{WG}	Within-class sum of squares
t	t-distributed variable
t_{df}	Student's t-distribution

$\mathcal{X}_i$	Sample set belonging to class i
$\mathcal{X}$	Sample set
α	Type-I error probability
β	Type-II error probability
γ	Parameter of Gaussian kernel
δ	Effect size of the population
η^2	Effect size
μ	Mean of the population
ν	Degree of freedom
ω^2	Effect size
Φ	Cumulative distribution function of the standard normal distribution
σ	Standard deviation of the population
σ^2	Variance of the population
σ^2_{BG}	Between-groups variance of the population
σ^2_T	Total variance of the population
σ^2_{WG}	Within-groups variance of the population
$\mathbf{\Theta}$	Parameter set
Ξ	Population set
Ξ_i	Population of class i

Recognition of Affect in Gait Patterns

c	Number of classes
C	Regularization parameter
$\mathbf{C}$	Covariance matrix
d	Dimension
f	Frequency
$f(\cdot)$	Decision function
$k(\cdot,\cdot)$	Kernel function
m	Dimension after mapping
$\mathbf{m}$	Total mean
$\mathbf{m}_i$	Mean of class i
N	Total number of samples
n_i	Number of samples in class i
P	Probability
P_E	Error rate
$P^\star_E$	Bayes error rate
$P(\cdot\|\cdot)$	Conditional probability
$p(\cdot)$	Probability density function
$p(\cdot\|\cdot)$	Conditional probability density function
$\mathbf{p}$	Eigenposture
P_E	Probability of error
$P^\star_E$	Bayes error rate
$\mathbf{S}_B$	Between-class scatter matrix
$\mathbf{S}_W$	Within-class scatter matrix

t	Time
T	Number of frames
$\mathbf{u}$	Eigenvector
$\boldsymbol{w}$	Normal vector of the hyperplane
$\mathbf{w}_k$	Description of walk k
α	Parameter
γ	Parameter
λ	Eigenvalue
$\boldsymbol{\mu}$	Mean of the population
Φ	Phase
$\boldsymbol{\phi}$	Map
σ	Standard deviation
$\boldsymbol{\Sigma}$	Covariance matrix of the population
ω	True state of nature
ξ	Slack variable

HMM for Recognition of Affect and Identity in Marker-Based Gait Analysis

$\boldsymbol{A}$	Transition probability matrix
a	Transition probability
$\boldsymbol{B}$	Observation probability matrix
b	Observation probability
$b(\boldsymbol{v})$	Probability density function for oberservation $\boldsymbol{v}$
$D(\cdot,\cdot)$	Distance metric
d	Number of observations
C	Number of classes
c	Class
c_t	Scaling factor
$\mathcal{E}$	Set of stances
$\boldsymbol{e}$	Stance
I	Number of states ω
K	Number of stances
N	Number of training sequences
$\mathcal{N}$	Standard normal distribution
$P(\cdot\|\cdot)$	Conditional probability
$\boldsymbol{S}$	Sample sequence
s	Deviation
$\boldsymbol{s}$	Sample
T	Sequence length
t	Time
v	Observation
$\boldsymbol{v}$	Observation vector
$\boldsymbol{V}$	Observation matrix
w	Weight
$\boldsymbol{w}$	Weight vector

α	Forward variable
β	Backward variable
Γ	Set of joint angles
γ	Probability
$\boldsymbol{\Theta}$	Parameters of HMM
$\boldsymbol{\pi}$	Initial state distribution
$\boldsymbol{\Sigma}$	Covariance matrix
σ^2	Variance
ω	True state of nature

Expressions of Emotions in Gait Patterns for Robots

F	Froude number
$F(\cdot, \cdot)$	F-distributed variable
g	Gravital acceleration
h	Step height
l	Link
l_h	Hip height
n	Number of walkers
p	p-Value
t	Step cadence
v	Velocity
w	Step width
y	Parameter
α	Joint
β	Joint
γ	Joint
η_p^2	Effect size (partial)

List of Figures

List of Tables

1 Introduction

Imagine, you are in a 22th century household - robots are helper for daily tasks such as cleaning, serving drinks and meals, and taking care for a person's needs - the door opens and a robot comes in. You would expect a friendly robot which serves your wishes, which communicates easily and intuitively with you, and which possibly makes you laugh. In social robotics, robots will cross the line from being task-oriented machines to machines interacting with their environment. To understand quickly and clearly the wishes of humans, verbal and non-verbal communication is essential in human-robot interaction. A lot of progress has been made in speech recognition since the early beginnings in the 80s. Yet, non-verbal communication is a new research field in robotics starting with the 21th century. Humans naturally express their intentions by verbal and non-verbal communication. Although verbal communication is more detailed and explicit, nonverbal communication gives an impression on the atmosphere and strongly relates to higher cognitive tasks such as decision making [123]. In the situation described above humans tend to transfer their knowledge about nonverbal communication to robots [130]. Hence, the robot should not only be able to understand human non-verbal behavior, but also behave itself in an appropriate and friendly manner. One way to achieve this behavior is that the robot behaves and signals its intentions similarly as humans do.

Non-verbal communication includes the recognition and the expression of task-relevant information, intentions, and emotions. Considering human-human interaction, the relevance of the latter term emotion is easily understandable. It is well known in psychology that emotions influence ones attention, motivation, memory, reasoning, and decisions [123, 133]. Yet, combining the terms emotion or affect with the terms computer or robot seems to be challenging at first glance. A computer or robot should fulfill dutifully its arithmetic problems and executing its tasks. Yet, when robots come into households or more general into the social domain, it would benefit from an entertaining character. This includes that the robot understands social behavior of the human and reacts appropriately. The robot would be capable to understand the emotional editing to task instruction, which could give information on the importance or difficulty. Furthermore, the robot would be able to reflect whether its own behavior was appropriate. Picard [123] introduces the term affective computing and gives several examples for what it is useful including social assistance and entertainment.

Affective computing is subdivided into recognition of a human's emotional state by means of machine learning algorithms, developing computational models for cognitive processes, and the design of emotive expressions for robots. Emotions can be expressed by several modalities, which are facial expressions, speech, physiology, gestures, and motions. Research in affective computing has focused predominantly on particular modalities which are well studied in psychology. However, a single modality is often not sufficient to estimate emotions reliably. For that purpose, multi-modal systems gain increasing importance.

Within this context, this work studies to what extend gait can be considered as a modality for affective computing.

1.1 Challenges

This interdisciplinary approach faces several challenges. First, introducing pattern recognition algorithms to gait analysis is an active research field in biomechanics and medicine. Second, dealing with emotions in machine learning rises special issues which are related to the broad and not unique definition of emotion. Third, modeling motion styles with parametric models is a sophisticated research area in social robotics. Each of these challenges are discussed more detailed in the following.

Pattern Recognition in Gait Analysis

Traditionally, inferential statistics is applied in biomechanical and medical gait analysis. Within the last decade, several approaches have been developed to apply techniques from machine learning to gait analysis [21, 22]. Inferential statistics and predictive statistics in machine learning differ in the way that the former analyzes how gait parameters are affected by various factors, and the latter investigates predicting causes of gait variations from observations. A detailed study whether these methods relate to each other and what conclusions can be drawn from inferential for predictive statistics and the other way round has not yet been accomplished. A study dedicated to this topic is beneficial for interdisciplinary research e.g. in biomechanics, affective computing, and bioinformatics.

Furthermore, gait databases are characterized by high dimensionality, temporal dependency, high variability, and nonlinearities. For these reasons, classification of gait patterns is a challenge for algorithms in machine learning. The review [21, 22] on techniques for marker-based gait analysis points out that several algorithms have been applied to gait analysis, yet a comparison of different methods is still an open research field.

Emotion as Classification Target in Machine Learning

A universal definition of the term emotion does not exist in psychology. Instead a broad interpretation and a number of emotion models exist. On account of this, several issues go along with developing pattern recognition algorithms which are specific for affective computing.

First, the ground truth of databases is questionable. Recorded data can be affected by lying and mixed emotions. This also influences the validity of judging the emotional states based on different modalities. It is not ensured that different modalities convey in general the same subset of emotions and that all modalities express the same emotional state at each point in time. Thus, mixed and blended emotions are an issue for multi-modal emotion recognition and specifying which emotions are best expressible and recognizable for a modality is essential. Given that psychology provides several models for categorization of emotions, it is beneficial to investigate which emotion model fits best to a modality.

Some bodily expressions of emotions are universal. However, a large part depends on the culture, on social norms, and on individual differences. It is expected that emotion expression varies among humans for these reasons. These variations additionally complicate automatic recognition.

Finally, facial expression and speech are well investigated in affective computing. Several review articles compare different algorithms and a number of gait databases have been recorded, some of them published [18, 118, 154]. For other modalities, especially the expression of emotions in body movements, a smaller number of studies exist and one database is public available [98]. In order to efficiently include body movements as modality in multi-modal emotion recognition systems, further studies are necessary which investigate the suitability of different motions, e.g. gait, in more detail. This provides a basis for multi-modal systems which go beyond speech and facial expressions e.g. in situations without an interaction or at distance.

Parametric Models for Motions Styles

Modeling different motion styles is beneficial for computer graphics, animation, and robotics. Instead of designing a limited number of motion styles, a parametric model has the advantage that a large number of expressions can be generated with a small number of setting values. Furthermore, developing models which are independent of the structure of the animated character or of the robot provides more general application of motion styles.

Yet, a coding system how emotions are encoded e.g. similar to the facial action coding system, does not exist for daily motions such as walking. Closest to such a description come the Laban features in dance theory [55]. To avoid dependency on the designer of the expressive motions, a challenging approach in this research field is to map human motion styles to virtual characters and robots via a mathematical model. The application to different body compositions is of special interest. Further aspects are whether expressability of virtual characters and robots differ and which emotions are better expressible in whole-body motions. Studying the expressiveness of robots facilitates non-verbal HRI in which the robot gives emotional feedback to its interaction partner.

The main challenge addressed in this thesis is to explore the suitability of the daily motion gait for emotion recognition and emotion expression in affective computing. This investigation faces the development of suitable algorithms in machine learning and interdisciplinary research combining computer science, psychology, and biomechanics. Possible application scenarios are non-verbal communication in HRI, monitoring of high-security areas, and entertainment.

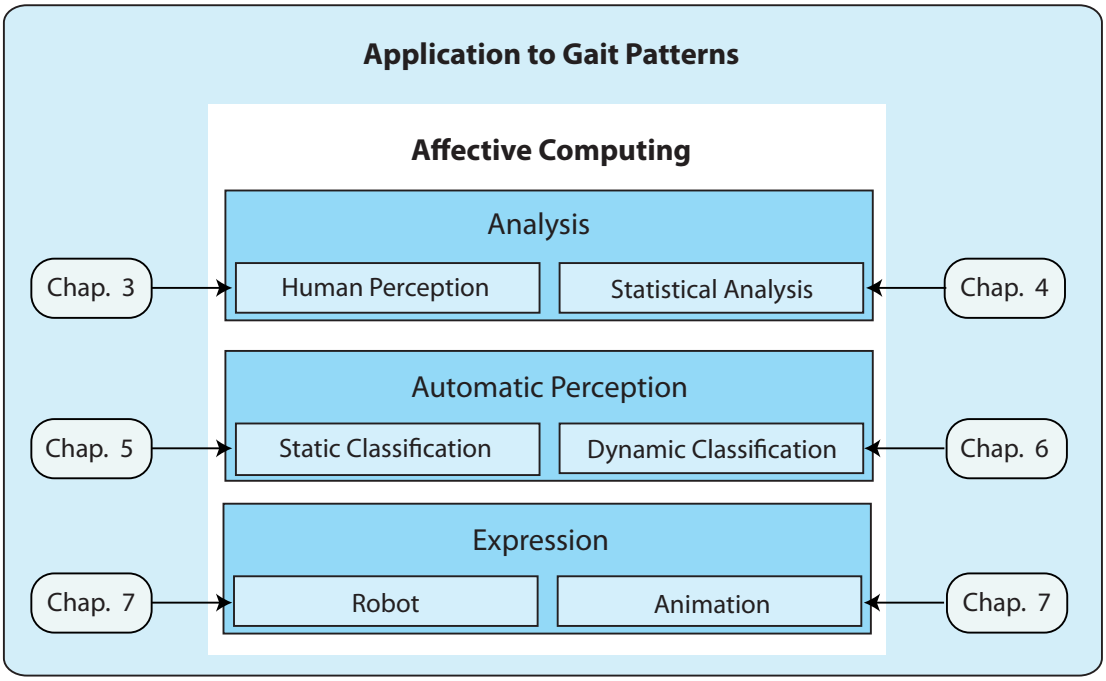


Fig. 1.1: Outline of the thesis.

1.2 Main Contributions and Outline of the Thesis

This thesis investigates the whole-body motion gait as modality for affective computing. It explores aspects on the human perception of emotions which are relevant for this application, addresses issues which are related to automatic emotion recognition in nonverbal HRI, and finally provides a mathematical concept for implementing expressive gait patterns on robots. Fig. 1.1 illustrates the outline of this thesis. Considering affective computing, three topics are of main interest when a modality is studied. These are analysis how humans express emotions via this modality, developing and evaluating pattern recognition algorithms for automatic recognition, and studying whether robots can express emotions utilizing this modality. These topics are investigated in the context of their application to gait patterns in the following chapters.

The *analysis of emotion expression and recognition in gait* is divided in two major topics. On one hand, it is studied how humans decode emotions from observing gait. In this context, several experiments are conducted and analyzed. Focus of interest is the recognition rate of humans, the comparison of two emotion models, visual gaze behavior during observing emotive gait patterns, and the interrelation of facial expressions and expressive walking styles. On the other hand, the expression of emotions in gait can be analyzed with inferential statistics. This method is traditionally applied in psychology and in biomechanics. Inferential statistics is compared with predictive statistics in order to draw conclusions whether quantitative relations between the two approaches exist. In doing so, similarities, differences, and mathematical relations between a selection of methods from inferential statistics and techniques from machine learning are elaborated. Particular focus lies on deducing classification rates from reported test statistics.

Based on the knowledge about human performance, pattern recognition algorithms are developed for *automatic recognition of emotions*. Techniques for dimension reduction and classification are investigated and compared. The latter is divided in static classification and dynamic classification which models the development of the gait trajectories over time. With regard to emotion recognition, two emotion models are compared as well as person-dependent versus inter-individual emotion recognition. Recognition rates are finally interpreted in the context of recognition rates for gender and identity.

The *expression of emotions in gait patterns* can be utilized for virtual character animation and robots. Expressive walking styles are developed for a walking robot. The walking styles are evaluated for a robot and its animation. In contrasted to related work, the walking styles are not designed by an expert. Instead, a model is derived from gait recordings.

The contributions to the state of the art are summarized separately for each topic hereinafter.

Analysis of Emotion Expression and Recognition in Gait

Chapter 3 starts with reviewing psychological studies about emotion expression and recognition in gait patterns. As a result, it can be concluded that emotions are expressible in gait patterns. Yet, only a small numbers of studies exists in comparison to other modalities. Therefore, several additional experimental studies on the human perception of emotions from gait patterns are conducted in the context of affective computing. They extend the state of the art with contributions about recognition rates for different emotion models and recognition rates for emotion recognition in comparison to gender and identity recognition. Furthermore, the visual gaze behavior during observing emotive walking styles is investigated. Even if no facial expressions are presented, human observers tend to observe especially the upper part of the body during an emotion recognition task. The results indicate that this is controlled by a top-down attention mechanism. Finally, if both facial expressions and expressive walking styles are combined, authenticity of the expressed emotion is higher, if both modalities express the same emotion. The studies provide 1) a reference to compare human performance with performance of machine learning algorithms, 2) valuable insights in the human perception mechanism of emotions, and 3) a motivation to develop expressive walking styles for virtual characters and robots to increase authentic emotion expression even if facial expressions may be available.

Part of the reviewed, psychological studies analyzes human gait with statistical inference [32, 106]. Furthermore, t-test or analysis of variance (ANOVA) are often applied in biomechanics to study the gait. Within the last decade, several approaches from machine learning have been introduced to gait analysis. This motivated to study how inferential and predictive statistics relate to each other in chapter 4. This work focuses on methods which are univariate and assume a Gaussian distribution underlying the data. For this, similarities, differences, and mathematical relations are elaborated. Results are exemplified for a psychological study on the embodiment of depression in gait. In doing so, classification rates of a linear Bayes classifier are derived from reported test statistics. For illustration, lower and upper bounds for the approximation are derived for the ANOVA. Furthermore, t-test and ANOVA are methods to select features in machine learning. This

approach provides a mathematical relation between the Bayes error rate and the criterion after which the features are sorted. Hence, values of interest in machine learning can be approximated from reported test statistics of psychological or biomechanical studies. This approach is of interest for interdisciplinary work in which researchers in medicine, biology, biomechanics, or psychology collaborate with computer scientists.

Recognition of Emotions in Gait with Machine Learning Algorithms

During the last decade, several approaches have been investigated to introduce methods from machine learning to gait analysis. The main application is therapeutic support in the research area of marker-based gait analysis. Yet, only one study has applied pattern recognition algorithms to recognize affect in gait [69]. Results of chapter 5 are 1) that recognition of emotion is enhanced if the identity of the walker is taken into consideration, e.g. if first the identity is recognized and then the affect, 2) differences in the affective dimensions arousal and dominance are better recognizable in gait than differences in pleasure with techniques from machine learning, and 3) the gender or the identity are easier to perceive from observing a person's walking style than affect.

In addition, this work contributes with a detailed analysis of different feature extraction and classification methods for marker-based gait analysis. Modeling gait trajectories by a combination of principal component analysis (PCA) and Fourier transformation does not outperform the simple approach to take minimum, mean, and maximum of each trajectory. Whether PCA or linear discriminant analysis achieves higher recognition rates depends on the rank of the within-class scatter matrices. Even though kernel methods have the advantage to take nonlinearities into consideration, recognition rates are lower for this application. The static classifiers nearest neighbor, Naive Bayes, and support vector machine are applied after feature extraction. If the feature extraction is efficient, the recognition rates varies slightly for the different classifiers. If the number of training samples is small, the feature vector is high-dimensional, and linear discriminant analysis is applied for dimension reduction, it is derived that the decision borders of a nearest neighbor classifier coincide with the decision borders of a hard-margin classifier and a support vector machine with a Gaussian kernel. Results on different techniques of machine learning for gait analysis have been published in [166, 167, 170, 171, 173, 175].

Chapter 6 approaches classifiers which incorporate the development over time of the gait trajectories. These are a minimum distance classifier for each stance and a continuous density hidden Markov model (CDHMM), which additionally models the transition from one gait stances to the next. In the context of affect and identity recognition, the result is that despite its good performance in vision-based gait analysis, it does not achieve higher recognition rates for marker-based gait analysis than static classifiers with efficient feature extraction.

Summing up, pattern recognition algorithms reach similar recognition rates as humans for affect recognition in gait. As gait is a highly individual motion pattern, recognition systems benefit if they take the identity of the walker into consideration. Even though dynamic classification models the transition between gait stances and is a promising concept, it does not necessarily outperform static classification with efficient feature extraction.

Expressive Gait Patterns for Robots

Artists or experts design expressive movements in related work. Chapter 7 follows a different approach. Characteristics of emotive gait patterns are derived from the recorded gait database and a model is developed. This model is applied to a walking robot, here a hexapod. Results of the evaluation are that observers recognize different emotions in the way the robot walks, especially differences in arousal and dominance. The expressiveness between animated and real robots differs only slightly. Results have been published in [172].

Thus, the walking style of the robot influences its appearance and perception by humans. This can be utilized in non-verbal HRI.

In conclusion, this work contributes with 1) insights on human perception of emotions in gait, 2) comparing inferential with predictive statistics, 3) developing, evaluating, and comparing pattern recognition algorithms for gait analysis, and 4) modeling expressive gait styles for walking robots. This work is assigned to the context of affective computing, yet especially the comparison of inferential and predictive statistics and the study on machine learning techniques for gait analysis find use in a wider range of applications.

2 Emotions in HCI/HRI

Humans interact socially with computers and robots [130]. To increase natural interaction, the integration of emotions in HCI/HRI is studied within the research field of affective computing. This includes automatic emotion recognition, synthesized emotion expression, and modeling the temporal behavior of emotions. In contrast to *mood*, *emotion* is characterized by short-term duration and is directed towards an external stimulus or reinforced event. Lazarus in [90] defines emotion as the combination of physiological disturbance, action tendencies, which are not necessarily acted out, and *affect*, which is the subjective experience during an emotion.

In human-human interaction, '*emotional intelligence* consists of the ability to recognize, express, and have emotions, coupled with the ability to regulate these emotions, harness them for constructive purposes, and skillfully handle the emotions of others' [125]. Social robotics is directed towards a similar ability of the robot to intelligently interact with humans. At this point, it differs from task-oriented robotics which focuses on intelligent manipulation within an environment.

The following sections provide background information on the concept of emotion in psychology, on current achievements in emotion recognition, and on emotion expressions for robots. It finally summarizes main applications and challenges in the research field of affective computing.

2.1 The Nature of Emotions

Within the last century, it has been shown that emotions influence human behavior. Too much emotion can hinder intelligent behavior whereas too little emotion leads to a lack in the ability to make rational and intelligent decisions in daily life [36]. This affects all cognitive processes such as attention, reasoning, decision making, motivation, memory and perception [123, 133]. For example, emotion influences attention such that emotional, in particular fear-relevant, events are detected more rapidly [41]. Furthermore, humans especially remember events that evoke emotions [41]. Predominantly, the effects of arousal on cognitive processes have been studied. High arousal facilitates short term forgetting and long term remembering [25, 80]. Also, high awareness of arousal enhances predictive judgment [41]. Performance for task accomplishment requires an optimal level of arousal [158]. Too little arousal facilitates a lethargic condition, however, too high arousal facilitates a hyperactive condition that can inhibit concentration.

Due to the diverse, versatile and elusive appearance of emotions in daily life, a universal definition of the term emotion does not exist. Instead, the term emotion is interpreted context-specific. Defining emotion for scientific analysis can be distinguished in three main directions, namely the categorical, dimensional and appraisal-based approach.

Fig. 2.1: P. Ekman's and W.V. Friesen's [46] intercultural studies resulted in the definition of the six basic emotions, which are anger, fear, joy, sadness, disgust, and surprise.

Emotions are classified as discrete affective states in the categorical approach, see Fig. 2.1. Analysis of facial expressions is often based on the six basic emotions anger, disgust, fear, joy, sadness, and surprise [46, 161]. As this basic emotions are pan-cultural, it is concluded that they are a product of evolution and not a result from culture-specific learning [46]. Discrete categories are not limited to the basic emotions, but they or a subgroup of them are studied most frequently in affective computing. This approach is supported by neurobiological findings showing that e.g. fear and disgust have different neural correlates [17].

The main disadvantage of discrete emotional states is that they cannot be merged and averaged in a meaningful way. To overcome this limitation, a set of basic dimensions has been proposed. The emotional state is described as a point in a space spanned by the basic dimensions. The independent and bipolar dimensions for the PAD model are pleasure, arousal and dominance [135], see Fig. 2.2. Pleasure is defined as how positive or negative an emotion is. Arousal corresponds to the level of activation, mental alertness, and physical activity - shortly the "call to action". Dominance is described as a feeling of control and influence over ones surroundings and others versus feeling controlled and influenced by situations and others. The dimensions pleasure and arousal are in accordance with the circumflex model of affect [134]. Personality, such as the characteristics sex, extroversion, neuroticism, and physical activity, influences emotional traits [104], e.g. men tend to be less aroused and more dominant than women, extroverts are behaviorally, more dominant, neuroticism influences emotional traits in various ways, and physical activity is associated with more dominant and more pleasant temperament. Mapping between discrete emotional states and the PAD model is investigated in [107, 111], e.g. the discrete categories such as happiness, amusement and contentment are related to high pleasure, whereas anger, fear and sadness are related to low pleasure. Both approaches are seen as alternatives. The PAD-model allows graduation of the emotional state, while discrete states are more related to general linguistic usage.

Due to its complexity, the appraisal-based approach is less used within the research field of affective computing. It finds more application as a psychological concept to explain how emotions develop, influence and are influenced in turn by the interaction with the environment [133, 141].

From neurobiology, it is known that mainly the amygdala, which is a small almond-shaped structure containing thirteen nuclei, plays a key role in processing social signals of emotion. Further parts in the brain, which are associated with emotions, are the prefrontal cortex, hypothalamus and anterior cingulate cortex [17, 35, 52, 133].

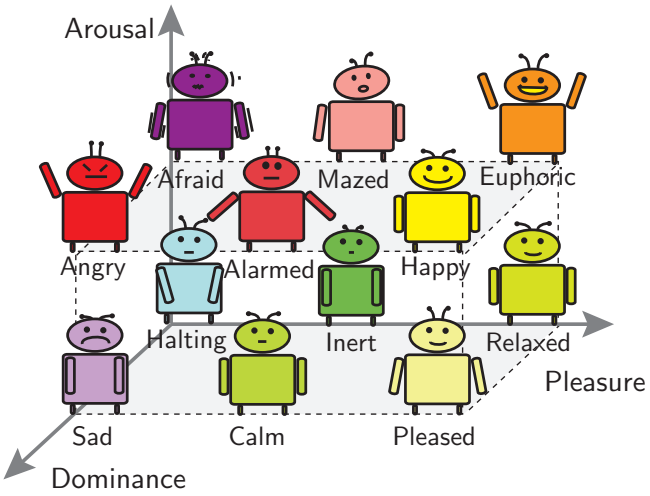


Fig. 2.2: The PAD model is a common alternative to the basic emotions in psychological research.

As the Latin origin of the word 'emotion'[1] implies, emotion is not only limited to personal feelings but it is also expressed in facial expressions, speech, gestures, and body motions. Studying the expression of emotions in man and animals is traced back to C. Darwin [37]. His contributions are a limited set of basic emotions and the finding that animal emotions are homologous for human emotions, which is a logical extension of the evolution theory. Traditionally, emotions have been seen in connection with bodily changes. In 1884, the James Lange theory came up that emotions are no more than the experience of bodily changes elicited by emotive stimuli [68]. Criticisms of this theory are that 1) bodily changes are typically to slow to generate emotions, 2) autonomic or peripheral changes elicited by injections or drugs do not determine which emotion is felt rather its magnitude, and 3) peripheral changes are not sufficiently distinct to differentiate the vast variety of emotions [19, 35, 133]. However recent experimental results draw attention on a modified version of the James Lange theory in which emotions are embodied such that body movements relate to the experience of emotional states e. g. the action smiling facilitates positive recognition of events providing scientific evidence for the proverb 'if you smile, the world seems to smile with you' [35, 116] .

2.2 Emotion Recognition in Machine Learning

The emotional feeling itself still remains hidden and is in its full range of diversity hardly locatable in e.g. brain structures, whereas it affects behavior and condition in a manifold way [17, 35]. These changes in behavior are observable for interaction partners and, in general, the environment. This refers to facial expressions, linguistic as well as acoustic characteristics of speech, body posture and motions. Physiological parameters such as heart rate, skin impedance, aspiration rate, and blood pressure are affected but without additional tools seldom visible to the environment. As facial expressions and speech dominate

[1] 'emovere' = to move out

during face-to-face interaction, these modalities are predominantly studied [56]. However, they cannot cover the full range of emotional states. Also, intended or non-intended antagonisms can hardly be recognized by observing a single modality. This motivates the development of multi-modal emotion recognition systems. In general, automatic emotion recognition faces the following challenges.

First, a *database* needs to be recorded containing different emotional expressions for the modality/ies under investigation. Emotional expressions can be either acted[2], elicited or spontaneous, where the latter one is preferable but often not accessible in experiments. Eliciting emotions during an experiment is an alternative if gathering data containing spontaneous emotions is too time consuming and hardly accomplishable due to ethical issues and technical difficulties. Common means are self-induction by recalling emotional experiences, watching pictures or videos with emotional content, listening to emotive sound or music clips, and eliciting emotions by execution of defined tasks in a controlled experimental setting. Posed emotional expressions represent an approximation of really felt emotions [165]. Recognition rates for spontaneous emotions from real-life scenarios are usually lower than for acted emotions, as actors tend to produce stereotypes or exaggerate expressive behavior.

As various *models* for the categorization of the term 'emotion' exist in psychology, see 2.1, the researcher needs to refer to one model during planning the study. Most approaches for automatic recognition investigate the categorical emotions, in particular the basic emotions or a subset of them. Fewer studies refer to emotional dimensions [56]. Nevertheless, recognition systems based on a dimensional model for emotions seem to be more appropriate for multi-modal emotion recognition because modalities differ in their intensity and distinctness to express various emotions. A single study investigates a framework using the appraisal-based approach [38]. In this study, the elicited emotional state is estimated with various techniques from machine learning under the assumption that the appraisal is known. However, estimation of the appraisal remains an open question.

Furthermore, recorded data needs to be labeled to train and evaluate the classification. *Ground truth* of the expressed emotional state is a critical aspect, as even self-assessment can be manipulated by lying. Believability and authenticity of emotional expressions is usually increased, if all modalities express the same state. Incongruence is correlated with lying. In this case, expressions of body movements seem to be more reliable than facial expressions because people do less bother to censor their body movement or physiology in daily life than facial expressions [45, 56] . Findings in nonverbal behavior support this view e.g. truthful expressions in speech differ from deceptive expressions in a lack of head movements and gestures [15, 56]. It should be noted that humans are highly sensitive to incongruence in emotive expressions and brain responses to these stimuli occur within 100ms [116]. As it is desired that an automatic recognition system performs at least as well as humans, the data is often labeled by human annotators. For this case, accuracy of 100% means that performance of the classification matches humans' skills. However, high inter-annotator agreement is often not obtained. How to distinguish for a certain case whether ground truth relates more to self-assessment, an annotator's rating, or neither of them, still remains an open question [56].

[2]Also referred to as 'deliberated' and 'posed' in related literature.

Pattern recognition is divided in *data preprocessing, dimension reduction and classification.* Data preprocessing is specific for each modality and delivers usually a high number of computable features. With a limited set of sample data, efficient dimension reduction is required to avoid the 'curse of dimensionality'. Applied techniques among others are Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Sequential Forward Search (SFS) and Sequential Backward Search (SBS). Common classifiers are Hidden Markov Models (HMM), Support Vector Machines (SVM), k-Nearest Neighbor (KNN), Bayesian Networks, Gaussian Mixture Models (GMM), decision trees, Artificial Neural Networks (ANN) and discriminant functions. Classification can be either applied to the data set of a single individual, later referred to as *person-dependent*, or to the samples containing trials of the person which is left out from the training samples, later referred to as *inter-individual*. The latter case generalizes the results to recognition of emotional states for unknown persons. As emotion expression varies among individuals, person-dependent recognition is more accurate than inter-individual recognition . Still, inter-individual recognition rates above 90% are reached for facial expressions and for emotive speech considering a small number of emotions [161].

Further challenges in pattern recognition especially relevant for emotion recognition are 1) the *baseline problem*, 2) *variations in duration* as well as *intensity* of emotions, 3) *automatic segmentation*, and 4) *context specificity* of expressions. The baseline, which represents neutral or in other words no emotional expression, remains nearly constant over time for facial expressions. However e.g. in physiology, the actual bodily condition influences strongly the baseline, even for a single individual. As the baseline differs highly between individuals for all modalities, this leads usually to lower inter-individual than person-dependent recognition.

Levenson suggests that the duration of an emotion ranges approximately between 0.5 and 4 seconds [93]. Too long measurement periods lead to missing emotional states or measuring multiple emotions. Currently, measurement windows vary between 2-6 seconds depending on the modality [56]. For recognition based on physiology, no study has been reported sufficient results for window sizes lower than 3 seconds yet [76]. Thus, short-term emotions could be easily missed by measuring physiology.

Furthermore, the recorded data is mainly pre-segmented such that the corpus contains separate emotive sequences [161]. However, emotions develop dynamically in daily life leading to varying intensity, duration and onset of specific emotional states. An online recognition system requires to be capable of handling ambiguity during the change of emotional states, knowledge about the different temporal structures of each modality and how to combine them, and as well a dynamical model of human emotions, which gives context information on when, how often, and which emotions are most probably expected to appear.

Besides, interpretation of an expression is often highly context specific, e.g. a twinkle could mean encouragement, irony, or only be a reaction to a fly in the eye. Distinctly and well expressed emotions occur more seldom a day than many subtle expressions. Currently, only the former one would be recognizable with state of the art techniques. The latter one requires a high number of classes, which represent the variety of emotions, and recognition algorithms which are capable of detecting subtle differences. This also supports the

Modality	Advantage	Disadvantage
Facial Expression	• Existence of Coding System (FACS) • Frequent occurrence during interaction • Basic emotions predominantly studied • Non-intrusive	• Easy to control deliberately
Speech	• Frequent occurrence during interaction • Basic emotions predominantly studied • Existence of a psychological framework for encoding affective states • Non-intrusive	• Not available if speaker is silent
Physiology	• Arousal correlated to autonomic nervous system • Harder to control deliberately	• Contradictory relations between pleasure and physiological measures • Intrusive • High influence of actual bodily condition
Gestures	• Recognition from distance • Non-intrusive • Existence of a prototype for basic gestures	• Culture-dependent and highly individual • Relatively unexplored in psychology
Motion and Posture	• Recognition from distance • No interaction required • Non-intrusive • Less susceptible to deliberate social editing	• Relatively unexplored in psychology • Might not have one-to-one relationship with specific emotions
Brain (EEG)	• Less controllable	• Only in psychological studies, not yet applied in HRI/HCI

Tab. 2.1: Comparison of modalities for emotion recognition.

combination of several modalities to increase the chance to detect subtle expressions and to facilitate their interpretation. This can be either accomplished by decision level fusion or feature level fusion, where the former one is preferable because it allows asynchronous processing for each modality, greater flexibility in modeling such that each model can be modeled separately, and adaptive channel weighting for each modality [56]. Feature level fusion would also increase the already high-dimensional feature space further, which is not desirable from the perspective of machine learning.

Currently, research on affect recognition concentrates predominantly on the modalities facial expression and speech. Excellent and extensive reviews for both modalities and combination of them can be found in [16, 18, 118, 119, 154, 161]. Recognition rates range between 70% and 98% depending on database, person-dependence, number of classes, and spontaneous or acted emotions. However, not all applications of affective computing can be covered by these two modalities. To enhance the range of applications and to improve accurateness of multi-modal emotion recognition systems, further modalities are studied regarding the emotions they convey best and their suitability for affective computing. Table 2.1 compares the advantages and disadvantage of modalities under investigation. In summary, facial expressions and speech are best studied both in psychology and engineering. Extensive research for the other modalities is still necessary, so that they are as suitable as speech or facial expression for recognition tasks. Furthermore, it is worth noting that not all modalities are capable to express with same intensity and distinctness the same emotions. This concludes that not only feature-level fusion and decision level fusion is required, but also a methodology which is capable to handle the diversity of emotions.

2.3 Emotion Expression

For a small number of applications in affective computing, e.g. automatic music selection, monitoring of the affective state of humans is sufficient. The major part of applications in affective computing refers to social interactions between a virtual agent or a robot and a human. Here, the virtual agent or robot expresses its own emotions. This is not only beneficial to establish a bidirectional nonverbal communication, but also to achieve a illusion of life as J. Bates points out in [4].

Design rules go back to traditional character animation [89, 149]. For believable agents and robots three rules are adopted [4]. First, only one, clearly defined emotional state should be expressed at a time. Second, the expression matches the emotional state of the character. And third, accentuation of the expressed emotion facilitates its recognition. Possible techniques therefor are foreshadowing, appropriate timing, exaggeration, and simultaneous actions.

Depending on the scenario, numerous approaches have been undertaken to implement emotional expressiveness for robots. Implementations reach from emotive sounds and tones, gestures and actions, to implementing parts of facial expressions such as eyes or mouth. The expressions are designed for each robot by artists, experts, or researchers. For facial expression, the facial action coding system (FACS) exists and provides a mathematical model for implementation of facial expressions. A small number of robotic heads exist which cover a wide range of facial expressions [13, 83]. This requires a sophisticated hard-

ware. To bypass this, Hashimoto et al. propose a curved surface display to build a robotic head which is capable to express emotions [60]. In a psychological study, Curio *et al.* find after-effects for the dynamic facial expressions disgust and happiness [33]. Viewing a facial expression for a longer time results in the tendency to classify the following expression as similar to the previous one.

Bethel and Murphy provide a review in [8] on non-verbal and non-facial affective expression for appearance-constrained robots. These robots lack expressive faces due to limitations caused by the application, the environment, or power and platform size. Therefor they recommend five main methods for affective expression in their correspondence which are body movement, posture, orientation, color, and sound. Furthermore, they introduce the concept of proximity zones to HRI which follows the theory of proxemics for human-human interaction in psychology. The distance between human and robot is divided in intimate $(0 - 0.46\text{m})$, personal $(0.46 - 1.22\text{m})$, and social $(1.22 - 3.66\text{m})$ proximity zones. Body movements and posture are suitable for the personal and social proximity zone. In the intimate proximity zone, the full body of the robot may not be visible depending on the size of the robot. Furthermore, the review points out that only a small number of research studies has focused on appearance-constrained robots for affective expressions, even though the proposed methods provide possibilities to equip task-oriented robots with affective expressions by small modifications at the existing robot hardware.

This motivates to further pursue research in HRI which provides a wide range of modalities by which a robot expresses affect in different proximity zones and for different requirements on the robot hardware.

2.4 Affective Computing

The term affective computing was coined by R. Picard in 1995 defining "computing that relates to, arises from, and deliberately influences emotion" [123, 124]. It considers giving emotional capabilities to computers. Motivation for this new research field came from enhancing machine learning algorithms by mimicking human cognitive capabilities. Human drives such as motivation, attention, and emotion play a central role for decision making, memory, perception, learning, creativity, and social interaction. This understanding has lead to a rethink of the role of emotions in computing. Within the last decade, an increasing number of research projects investigate how emotions can automatically be recognized, how emotional drives can be integrated in cognitive architectures, and how synthetic emotion expressions can be designed. Fig. 2.3 illustrates the increase of publications regarding affective computing within the last decade.

Applications for affective computing are

- *Monitoring*: Monitoring the emotional state in daily life provides objective measures for medical treatment of e.g. depression and autism [125]. In road traffic, monitoring emotions, e.g. anger, and attentiveness provides input for regulations to increase driving safety [74]. In working environments, the emotion fatigue is of special interest to avoid maloperations of machines [173]. Work cycles can be adapted to the current condition of the worker. Furthermore, the way a robot executes its motion, influences

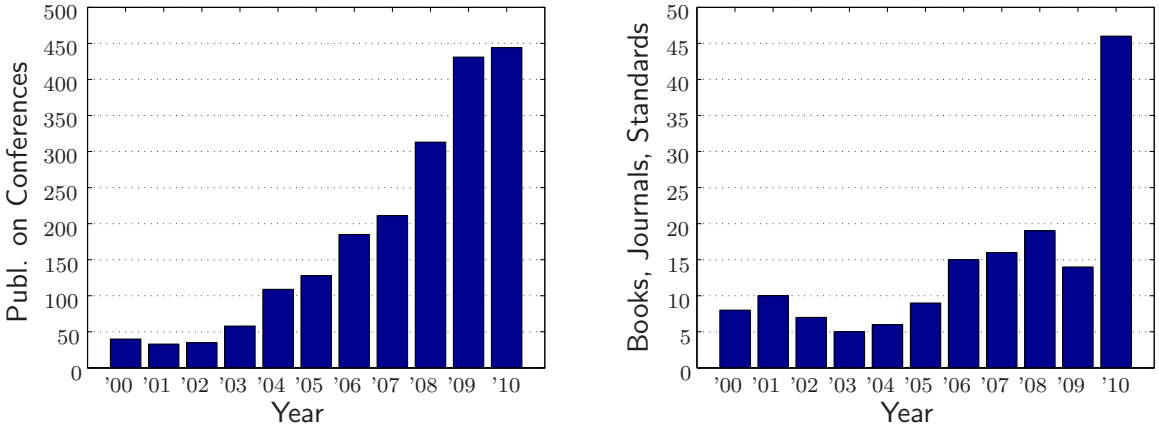


Fig. 2.3: Number of publications in conference proceedings and in books, journals or standards published by IEEE (provided by an IEEExplorer search with the keywords 'Emotion', 'Affective Computing' or 'Affective State' in Abstract, 29.07.2011).

a human's affective state in HRI. Robotic motions can be optimized to increase safety and decrease perceived anxiety of workers [84, 85].

- *Social Assistance*: Whether or not computers and robots have emotions, humans interact socially with them. Assistance systems such as guidance or household robots gain from learning personal preferences of their interaction partners, the sense when to interrupt, and leading small talk [123]. Another application is the automatic music selector which selects music depending on the emotional state of the listener [123]. A sophisticated, computer-aided learning software interacts emotionally to keep the user interested and avoid boredom [123].

- *Entertainment*: In 2007, a special exhibition on 'History and Presence of Human Machine Interaction' in Berlin was devoted to the history and development of robots. In fiction, a robot already forms its own identity including the sense for emotions. This idea has been adopted by the film industry and in the broader sense by art, e.g. the interactive robot theater [12]. The transfer of the design rules for believable characters from story telling and character animation leads to more entertaining robots and software agents. One of the primary means is to integrate emotions [4].

Summing up, studying automatic recognition of emotions, computational models for cognitive processes including emotions, and synthesized generation of emotional expressions has established as a serious research field. Potential concerns are that 1) people may expect human-like intelligence, understanding, and actions from emotionally reacting robots and software agents, 2) ethical issues regarding the assurance of a human's privacy and autonomy of decision making, and 3) the occurrence of unpredictable behavior.

3 Perception of Emotions from Gait Patterns

To study how humans perceive and express emotions is essential in affective computing. On one hand, results on human perception serve as a reference for automatic recognition and can give valuable input for model-based machine learning algorithms. On the other hand, these studies can be used to derive guidelines how virtual characters or robots can convey emotions.

In affective computing, the expression of emotions is well studied for speech and facial expressions. Yet, other modalities such as physiology, posture, and motions are less investigated. Decoding and encoding of emotions from walking is a relatively unexplored field in psychology, and, similar to body movements, it is assumed that a one-to-one relationship with specific emotions might not exist. This motivates to first analyze aspects of human perception which are later relevant for developing machine learning algorithms. Advantages of gait as a modality for emotion recognition are that it provides a means to detect emotions at distance, it is less susceptible to deliberate social editing, it does not require an interaction, and the observation is non-intrusive in comparison to physiological parameters.

Reviewing publications on psychological studies indicates that emotions are expressed in body movements, yet only a few studies considered explicitly gait. This motivates to analyze two gait databases in more detail. Focus lies on 1) estimation of the performance how well humans recognize emotions from gait patterns and which emotion model is most suitable, 2) visual behavior during the observation of gait patterns and during decoding emotional states, and 3) how the interrelation between different facial expressions and different emotional gait patterns influences the perception of emotions.

This chapter starts with an overview on psychological studies which refer to encoding and decoding of emotions from gait. Thereby, literature provides only recognition rates for categories of emotions. One contribution of this chapter is the determination of recognition rates for the dimensional emotion model so that both models can be compared. The obtained results show that specific emotions are encoded in the way a human walks for both models, yet emotions are harder to retrieve from a person's walking style than gender or identity. Human performance is compared to recognition rates of machine learning algorithms later in chapter 5 and 6. Furthermore, a descriptive study of the visual gaze behavior gives useful insights where human observers look during watching emotive gait patterns. The observation of the upper body is more relevant than observation of the lower body during an emotion recognition tasks. This is utilized in chapter 5 in which a subset of joint angles is analyzed for automatic recognition. The main goal of conducting a study in which facial expressions are combined with emotive gait is to investigate how relevant emotive gait patterns are if the face expresses emotions. The result is that congruency increases believability and reduces recognition time. This motivates to model the expression of emotions in gait for robotic applications in chapter 7.

The remainder of this chapter is organized as follows. Section 3.1 summarizes related work in psychology about encoding and decoding of emotions from gait patterns. Section 3.2 gives a description of the gait databases on which this work relies. The following section 3.3 investigates the human skill to recognize emotions in gait patterns. Recognition performance, gaze behavior, and the interrelation between walking styles and facial expressions are of special interest. Section 3.4 summarizes the results in the context of the current state of the art and section 3.5 ends with a discussion on the limitations.

3.1 Emotions in Gait

Various psychological studies indicate that humans are not only capable to recognize the intended action [14], but also gender [82], identity [34] and even emotions [14, 31, 40, 126] from body movements. The majority of psychological and behavioral studies on emotion recognition is based on categorical emotions. The categories happy, sad, and angry are more distinctive in motion than categories such as pride or disgust [31, 110]. Analysis of the arm movements drinking and knocking shows that the emotional states used in Pollick *et al.'s* [126] study are aligned with the arousal-pleasure space, if human judgments are taken as a basis. Furthermore, arousal is highly correlated with velocity, acceleration, and jerk of the movement.

Evidence exists in psychology that emotions can be expressed in walking and recognized by human observers. In 1987, Montepare *et al.'s* [110] psychological study indicates that observers can identify emotions from variations in walking styles. Furthermore, the emotions sadness and anger are easier to recognize than pride for human observers. Also, Michalak *et al.'s* [106] study supports that sadness and depression are embodied in the way people walk. Similar to facial expressions, emotions are expressed more intensely on the left side of the body during walking, which leads to a lateral asymmetry of bodily emotion expression [132, 137]. Crane and Gross [32] show that bodily expression of felt and recognized emotions are associated with emotion-specific changes in gait kinematics and not solely depend on gesticulatory behavior. They identify velocity, cadence, head orientation, shoulder, and elbow range of motion as significant parameters which are affected by emotions. Especially, the perception of fear is facilitated if the walker displaying fear is female due to similar kinematics for fearful gait and female specific gait [59]. Roether *et al.* [131] indicate that kinematic parameters which are critical for perception of emotion closely match features which correspond to changes in motor behavior. Although movement speed influences the perception of emotions, their study provides evidence of additional, emotion-specific features in gait that cannot be explained by variations in gait speed alone. Furthermore, they find no significant differences between recorded motions of lay-actors and novices, although the group with lay-acting experience reported less inhibition during the recording of emotional movements than did the novices. Tab. 3.1 summarizes the reported classification rates for emotion recognition.

Besides these findings, Heberlein *et al.* [62] suggest that the perception of affect and personality traits is accomplished by partially dissociable neural systems processing gait patterns in the brain. In current neurobiological models, emotional information is processed by a network involving both cortical and subcortical structures. A phenomenon called

Study	Year	Emotions	Visualization	Classification
Roether et al. [132]	2008	Neutral, anger, happy, sad	Unspecified	88%
Roether et al. [131]	2009	Anger, happy, sad, fear	Animated puppet	78%
Crane and Gross [32]	2007	Anger, content, joy, neutral, sad	Video	67%

Tab. 3.1: Literature survey on recognition of emotions in walking.

affective blindsight is the ability to process emotional signals in the absence of normal vision. This has been observed not only for facial expressions but also for emotional body language. This supports the brain's ability to process emotional body language unconsciously without reliance on the primary visual cortex [53].

One can summarize that psychological studies provide evidence that emotions are expressed in gait and humans can decode different emotions from observing gait patterns. Yet, only categories of emotions have been considered in the psychological studies. The dimensional PAD model, which is also suitable for technical applications, has not yet been investigated for different walking styles. Furthermore, the related studies have identified parameters in gait which are relevant for emotion expression. Still, little is known which body parts are preferably observed during an emotion recognition task and if the gaze behavior differs whether the emotional state should be recognized or not. Additionally, the gait has only been investigated as a single modality. How emotions are perceived if also e.g. facial expressions are visible and what effect different walking styles have in this case has not yet been analyzed. These aspects are studied within this chapter. Results are based on two gait databases which are described in the following.

3.2 Gait Databases

In order to study human perception and later on pattern recognition algorithms, this work investigates two motion databases which are the Emotive Motion Library [98] and the Munich Database [167]. Both databases were recorded with a marker-based optical tracking system. The Emotive Motion Library has been recorded with the intention to study the effect of emotions on different motions. The Munich Database focuses especially on the motion gait and has been recorded for this project with the purpose to investigate inter-individual versus person-dependent recognition. Tab. 3.2 lists the properties of each database which are relevant for estimation of human recognition performance and automatic recognition performance. These are recorded data, number of participants, recorded motions, recorded emotions, and number of repetition of each emotional state. The following description of each database goes into detail about the recording procedure, emotion elicitation, and recorded data.

Property	Emotive Motion Library	Munich Database
Recorded data	Marker positions	Marker positions, joint angles, joint centers
# Participants	15 female, 15 male	13 male
Motion	Walking, knocking, lifting, throwing	Walking
Emotion	Angry, happy, neutral, sad	Angry, happy, neutral, sad; low, medium and high expression of arousal, pleasure and dominance
# Repetitions	1	10

Tab. 3.2: Comparison of emotive motion databases.

3.2.1 Emotive Motion Library

F. Pollick [98] together with his team at the University of Glasgow, Scotland, recorded a motion capture library for computational and behavioral analysis of movement variability, of human character animation, and of how identity, gender, and emotion are encoded and decoded from human movement.

For the database, 30 non-professional actors performed walking, knocking, lifting, and throwing actions while they were recorded with a Falcon Analog optical motion capture system (60 Hz). All actions are performed during expressing one of the emotions angry, happy, neutral, and sad. These emotions were chosen because they represent emotional states which may last for an extended period of time. A story for each emotion was told to the participants, in which they had to imagine themselves in a specific situation. The position over time of 33 passive markers, which were attached to the skin of the subjects, is available in .csm format. The data has been downloaded from paco.psy.gla.ac.uk/data.php .

Participants walked in a triangle for 25 seconds to record longer sequences of emotive gait patterns. For estimation of human performance, 12 seconds of each gait recording is used. For automatic recognition, the triangle walk complicates recognition of emotions and may cause extraction of inappropriate features. Thus, sections of straight walking are extracted from the data sets, each with a length of 1.65 seconds, which is equal to 99 frames. The algorithm for finding this section searches for straight movement of the marker close to the bury center. The position of 33 markers of 29 participants within this time window is used later on for feature extraction and classification. Marker positions are illustrated in Fig. 3.1.

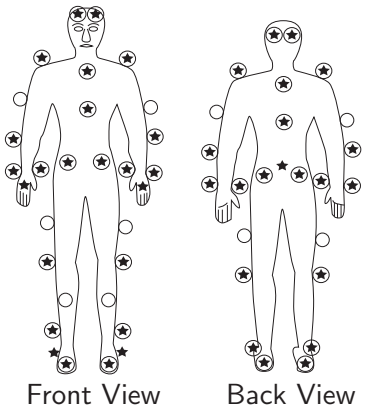


Fig. 3.1: Position of the markers during optical motion tracking for the Emotive Motion Library ($\star$) and the Munich Database ($\circ$).

3.2.2 Munich Database

A gait database has been recorded at TU München in collaboration with W. Seiberl and A. Schwirtz, Institute of Biomechanics in Sports, TU München, with the purpose to gain a larger understanding of the influence of emotions on gait patterns and its application to recognition of emotions [167]. Design of the database focused on inter-individual versus person-dependent analysis and the use of affective dimensions versus categorical emotional states.

The gait of 13 male non-professional actors (mean age: 25.8) was recorded with a VICON optical tracking system (240 Hz). As illustrated in Fig. 3.1, 35 passive markers were affixed to the participant's skin, where anatomic points define the marker positions. Participants were asked to feel angry, happy, neutral or sad, and to imagine a situation in which they feel a particular affect. A story for each emotion common for all participants was not told, because interpretation of the emotional content of such stories is subjective. Each recording contains at least two strides and each trial was repeated 10 times. Participants walked straight in the laboratory hall. In addition, extremes of the dimensions of the pleasure-arousal-dominance (PAD) model were recorded, so that the gait database contains also 10 walks of each participant expressing the affective states displeased, content, bored, excited, obedient and dominant.

Due to a highly artificial environment and frequent repetition of each affective state, successful elicitation of affect is challenging and it has been decided to pose affect. Although actors tend to produce stereotypes or to exaggerate expressive behavior, some evidence exists in literature that posed affective expressions represent an approximation to really felt affective expressions [165].

Based on the Plug-in-Gait Model, the VICON software provides marker positions, joint centers, and joint angles over time for further data analysis [152]. Depending on walking speed and stride length, each recording contains between a single stride and up to seven strides. A single stride is extracted from each recording for consistency. Overall, the

database contains 520 strides for analyzing categorical emotional states and 780 strides for analyzing affective dimensions.

3.3 Human Perception

Motivated from psychological evidence that emotions are expressed in body movements, quality of expressed emotions in both databases is judged by human observers. Furthermore, it is analyzed whether the PAD model is also suitable besides categories of emotions for studying emotions in gait patterns. As still little is known on the perception of emotions from movement, further experiments deal with the visual gaze behavior during an emotion recognition task and the interaction of facial expressions and walking style.

3.3.1 Recognition Performance

The performance how well humans can recognize emotions from observing gait is investigated in the following. In the experiments, human participants rated different walking styles concerning the emotional state of a walker. How well emotions are recognizable in comparison with gender or identity is analyzed utilizing the Emotive Motion Library. Analysis of the Munich Database concentrates on the comparison of two emotion models.

Experiment I: Emotion Recognition versus Gender Recognition and Identification

In 1973, G. Johansson showed that human movement can be displayed by illuminating only the joints of the body as light points on a black background [70]. This technique is called *Johansson display* or Point-Light-Walker (PLW). As the Emotive Motion Library provides reliable measurements only for marker positions, PLWs visualize the marker positions over time for the evaluation. Overall a number of 116 (4 emotions $\times$ 29 walkers) PLW videos were generated each 12 seconds long (15 frames/second, resolution of 560 $\times$ 420 pixels). Each video starts with the calibration position of the walker followed by approximately 10-11 seconds walking in a triangle.

A number of 18 participants (11 female, 7 male, age: 24.2 ± 3.2) took part in the experiment. In the first part, the participants rated the emotional expression of the walker and if the displayed walker appears to be either male or female. In the second part, the participants assigned the identity to each presented PLW. To avoid false ratings caused by boredom, the maximum duration for the experiment was set to 30 minutes. Therefore, the set of PLW videos was divided in six subsets, each containing the videos of 5 walkers. Thus, each participant rated five walkers, in doing so each walker was rated by three participants[1]. For recognition of the identity, a participant first watched the four gaits of each of the five walkers in the subset and was allowed to make notes on the walking style of each walker. Then, the PLWs were shown again in a randomized order and participants were asked to identify the walker of each video.

Tab. 3.3 summarizes the recognition rates for identity, gender and emotion. With an average recognition rate $\bar{R}$ of 74%, participants classified the identity best. In comparison

[1]Walker 25 was rated by 6 participants to allow an equal number of PLWs in all subset.

	Expressed Emotion				$\bar{R}$	Chance Level C	Extra Success
	Angry	Happy	Neutral	Sad			
ID	79%	77%	67%	76%	74%	20%	68%
Gender	61%	61%	53%	57%	58%	50%	16%
Emotion	54%	43%	54%	85%	59%	25%	45%

Tab. 3.3: Comparison of identification, gender, and emotion recognition.

Expressed Emotion	Recognition				# Trials (participants×walkers)	$\bar{R}$
	Angry	Happy	Neutral	Sad		
Angry	47	20	19	1	3×29	54%
Happy	7	37	37	6	3×29	43%
Neutral	5	12	47	23	3×29	54%
Sad	3	0	10	74	3×29	85%
Overall	62	69	113	104		59%

Tab. 3.4: Confusion matrix for emotion recognition for the Emotive Motion Library.

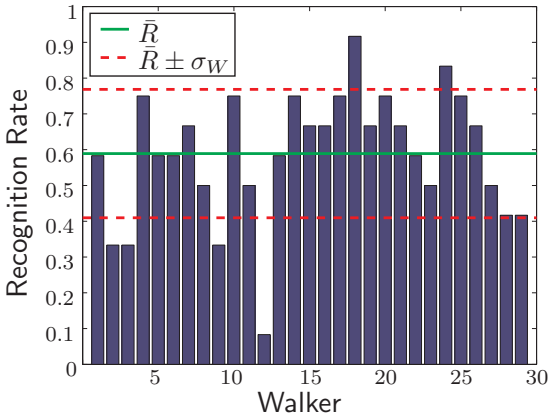


Fig. 3.2: Expressiveness of the walkers differ with $\sigma_W = .18$.

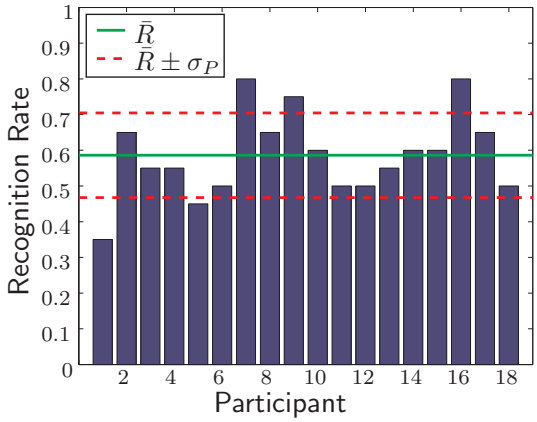
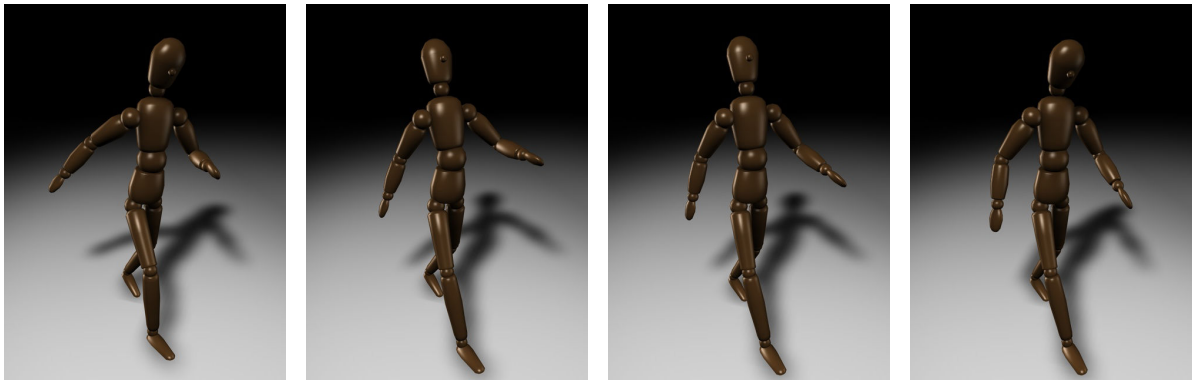


Fig. 3.3: Ratings of the participants differ with $\sigma_P = .12$.

to neutral walking, a walker is easier to identify if he/she expresses an emotion. Extra success $\frac{\bar{R}-C}{100\%-C}$ comparing to random predictor is 68% [2]. Even though the skeleton of men and women differ, recognition of gender is with 58% only slightly above chance level $C = 50\%$. Still, participants misclassified the attribute gender less if the walker expresses an emotion comparing to neutral walking. A comparable study reports that gender is 63% correctly classified in walking using PLWs [82]. Average recognition rate $\bar{R}$ of 59% for emotions is in the range of the recognition rate for gender, however chance level is 25%.

[2]The chance level C is 20% in this case, because only a selection of five walkers was shown to each participant. This limitation was set with respect to the human capacity for processing information [108].



(a) (b) (c) (d)

Fig. 3.4: Captured joint angles are mapped to an abstract puppet. The rendered animations have been evaluated by human observers. The snapshots show one walker expressing the emotions a) angry, b) happy, c) neutral, and d) sad.

Hence, the extra success of 45% is more than twice the extra success for recognition of gender. Tab. 3.4 shows the confusion matrix for emotion recognition. As each walker was rated by three participants, a total of 87 ratings is available for each emotion. Mostly, emotive gait styles are confused with neutral walking. This leads to the conclusion that the expressiveness of the walking styles has not always been sufficient to be perceived by an observer. Usually sad walking is the slowest gait pattern and is least confused. Furthermore, sad walking is characterized by a forward bend of the upper body, the neck, and the head.

Recognition performance differs between participants and also the expressiveness differs between the walkers, see Fig. 3.2 and 3.3. Differences in the expressiveness of the walkers is traced back to individual walking styles, individual expression of emotions and individual interpretation of the emotional content of the stories which have been told for emotion elicitation.

Experiment II: Categorical Emotions versus PAD-Model

Expression and recognition of categorical emotions versus different levels of expression on the affective dimensions pleasure, arousal, and dominance is analyzed utilizing the Munich Database. In contrast to the previous study, a graphical animation of an abstract puppet is used for visualization, which has been designed as part of a supervised bachelor thesis [182]. The abstract puppet provides a more realistic visualization of the gait patterns and still facilitates that human observers rate the expressed emotion only based on kinematic parameters without influences of physique or facial expression of the walkers on the evaluation. One gait trial of each emotion for each walker has been mapped to the animated puppet. Fourier Transform is applied to retrieve a parametric description of the captured joint angles $\varphi_i(t)$ over time, i.e.,

$$\varphi_i(t) \approx \sum_{j=1}^{10} A_j sin(2\pi f_j + \Phi_j) \tag{3.1}$$

Expressed Emotion	Recognition Angry	Happy	Neutral	Sad	# Trials (participants×walkers)	$\bar{R}$
Angry	112	34	34	15	15×13	57%
Happy	27	104	56	8	15×13	53%
Neutral	5	25	135	30	15×13	69%
Sad	7	6	40	142	15×13	73%
Overall	151	169	265	195		63%

Tab. 3.5: Confusion matrix for emotion recognition for the Munich Database.

where the ten frequencies f_j with highest absolute amplitude A_j and associated phase shifts Φ_j achieve a good approximation of the recorded angle $\varphi_i(t)$. The duration of a single stride is too short for human judgment; therefore, the individual gait patterns are extrapolated using the parametric Eq. 3.1, with each rendered video lasting 7s. Fig. 3.4 shows snapshots of the rendered movies for the emotions angry, happy, neutral, and sad. 30 participants (12 female, 18 male, age: 25.5 ± 3.5) took part in the experiment. To avoid boredom during the experiment, the upper limit for duration of the experiment was set to 30 minutes. Hence, animations were presented to two different groups: A and B. Group A watched all animations of walkers 1-6 and group B watched all animations of walkers 7-13.

Categorical Emotions: First participants rated the animated puppets that express four emotion categories. The average recognition rate $\bar{R}$ for four emotion categories is 63%, at this human observers tend to recognize, in particular, the emotion sad best, see Tab. 3.5. All recognition rates are approximately in the range of recognition rates of the previous study, see Tab. 3.3 for comparison.

Furthermore, participants were asked to rate, on a five-item Likert scale, how difficult it had been to estimate each of the emotions. A one-way repeated-measures analysis of variance (ANOVA) was used to test for statistically significant differences in the degree of difficulty across the four emotions. The degree of difficulty in estimating emotions significantly differs: $F_{3,87} = 2.91$, $p = .04$, $\eta_P^2 = 0.09$. Pairwise comparison indicates that the affective state sad (mean = 2.47) is perceived to be easier recognizable than the affective state neutral (mean = 3.23).

PAD Model: In the subsequent experiment, participants rated the animated walkers who expressed either low, medium, or high levels of pleasure, arousal, or dominance. Participants rated the level of the expressed affective dimension on a five-item Likert scale, i.e. corresponding row of the self-assessment manikin (SAM) questionnaire [88]. Accuracy for all three levels on each dimension is above chance. Mean accuracy for recognition of three values on the dimensions pleasure, arousal, or dominance is 55%, 61%, and 62%, respectively. Tab. 3.6 shows the confusion matrix for each dimension. Participants rated on a five-item Likert scale; therefore, the two lower and the two upper ratings were combined as 'low' and 'high', respectively. This explains the lower recognition rates for neutral gaits in comparison to high or low expressions of pleasure, arousal or dominance. The results indicate that different levels of arousal and dominance are better recognizable than different

Expressed Level	Pleasure				Arousal				Dominance			
	L	N	H	$\bar{\mathbf{R}}$	L	N	H	$\bar{\mathbf{R}}$	L	N	H	$\bar{\mathbf{R}}$
Low (L)	129	38	28	66%	153	37	5	78%	138	38	19	71 %
Neutral (N)	36	89	70	46%	78	62	55	32%	23	78	94	40 %
High (H)	41	59	104	53%	15	37	143	73%	13	36	146	75%
Overall				55%				61%				62%

Tab. 3.6: Confusion matrix for each affective dimension of the Munich Database.

Dimension	Effect	Group	df	df(Error)	F	p	η_p^2
Pleasure	Expression	A*	2	28	48.76	.00	.78
		B*	2	28	166.73	.00	.92
	Walker	A	5	70	1.77	.13	
		B*	6	84	23.86	.00	.63
	Interaction	A*	10	140	21.16	.00	.60
		B*	12	168	24.64	.00	.64
Arousal	Expression	A*	2	28	283.94	.00	.95
		B*	2	28	428.56	.00	.97
	Walker	A*	5	70	26.76	.00	.66
		B*	6	84	29.71	.00	.69
	Interaction	A*	10	140	23.40	.00	.63
		B*	12	168	17.56	.00	.56
Dominance	Expression	A*	1.26	17.57	65.51	.00	.82
		B*	2	28	513.85	.00	.97
	Walker	A*	5	70	21.12	.00	.60
		B*	6	84	11.91	.00	.46
	Interaction	A*	10	140	4.06	.00	.23
		B*	5.10	71.46	5.50	.00	.28

Tab. 3.7: Statistics for expression of affective dimensions in walking.

levels of pleasure in gait patterns.

A two-way repeated-measures ANOVA with the within-subject factors 'level of expressed affective dimension' and 'identity of walker' has been conducted separately for each dimension pleasure, arousal, and dominance. Main interest is to investigate if the participants' ratings differ significantly due to different affective expressions of the walkers. The following statistical analysis is conducted separately for group A and B. Mauchly's test indicates that the assumption of sphericity is violated for the independent variable dominance, $\chi_A^2(2) = 11.718$, $p < .05$ in group A and the interaction between dominance and identity,

$\chi^2_B(77) = 128.964$, $p < .05$ in group B. Degrees of freedom are corrected using Greenhouse-Geisser estimates of sphericity in these cases ($\epsilon = .425$ for the interaction effect between walker and dominance and $\epsilon = .627$ for the main effect dominance). No correction is necessary for the other cases.

All effects are reported as statistically significant at $p < .05$. Except for the effect walker on the dimension pleasure for group A, all tests are statistically significant and marked with a * after the group in Tab. 3.7. Tab. 3.7 lists level of significance and the effect size partial eta-squared η_p^2 for the tests of within-subject effects. Level of expression is either low, neutral or high. Perceived observations differ statistically significantly for level of expression on the dimensions pleasure, arousal, and dominance. Although, identity of the walker and interaction effects between walker and level of expression influence the ratings, the effect size indicates that level of expression explains most variance in the subjects' ratings. Hence, human observers are capable to distinguish levels of affective expression in the gait patterns of the data base. This perception relays more on a walker's current expression than on a walker's individual gait given by his physique. Subsequent pairwise comparisons for each dimension pleasure, arousal and dominance support this conclusion.

Finally, participants were asked to rate the degree of difficulty in estimating different levels of pleasure, arousal or dominance on a five-item Likert scale. In this case, one-way repeated-measures ANOVA indicates that there are differences in estimating different levels of either pleasure, arousal or dominance, $F(2, 58) = 7.09$, $p = .00$, $\eta^2 = 0.20$. The following pairwise comparison shows that pleasure is harder to estimate than different levels of arousal. Underlying reasons are that different levels of pleasure are harder to retrieve from gait patterns, are harder to express in walking, or a combination of both.

In conclusion, analysis of the recognition of affective dimension in gait patterns shows that different levels of arousal and dominance are better recognizable for human observers than different levels of pleasure. Furthermore, this explains the frequent confusion of angry and happy gaits because they share a similar level of arousal and differ mainly in pleasure. Based on human judgment of expressive gait patterns, the PAD model and categorical emotions are comparable approaches for further analysis.

3.3.2 Visual Gaze Behavior during an Emotion Recognition Task

In comparison to facial expressions little is known about the perception of emotions from body postures and movements; therefore, the question raises if specific parts of the body are especially observed during decoding of emotions from whole-body motions and if yes, which ones. It is assumed that the upper part of the body is more important than the lower part.

In connection with the second experiment, participants were asked to rate if specific body parts express in their opinion emotions during walking. Response options were 'yes', 'no', and 'do not know'. If the participants rated 'yes', they were further asked to specify the emotions expressed by that body part. Results are summarized in Tab. 3.8. The majority of participants found that the motion of the arms and legs, and the posture of the head express emotions during walking. From the ratings, it is further concluded that the motion of the arms and the posture of the head is slightly preferred for emotional judgment in comparison to the motion of the legs. Furthermore, the participants usually associated

Body Part	Emotional Expressiveness		
	Yes	No	Do not know
Motion of Arms	28	1	1
Motion of Legs	22	4	4
Posture of Head	29	0	1

Tab. 3.8: Results of a survey on emotional expressiveness of different body parts during walking (30 participants).

the affective dimensions arousal and dominance with leg and arm motion, whereas the posture of the head was found to be more expressive for happiness, sadness and dominance.

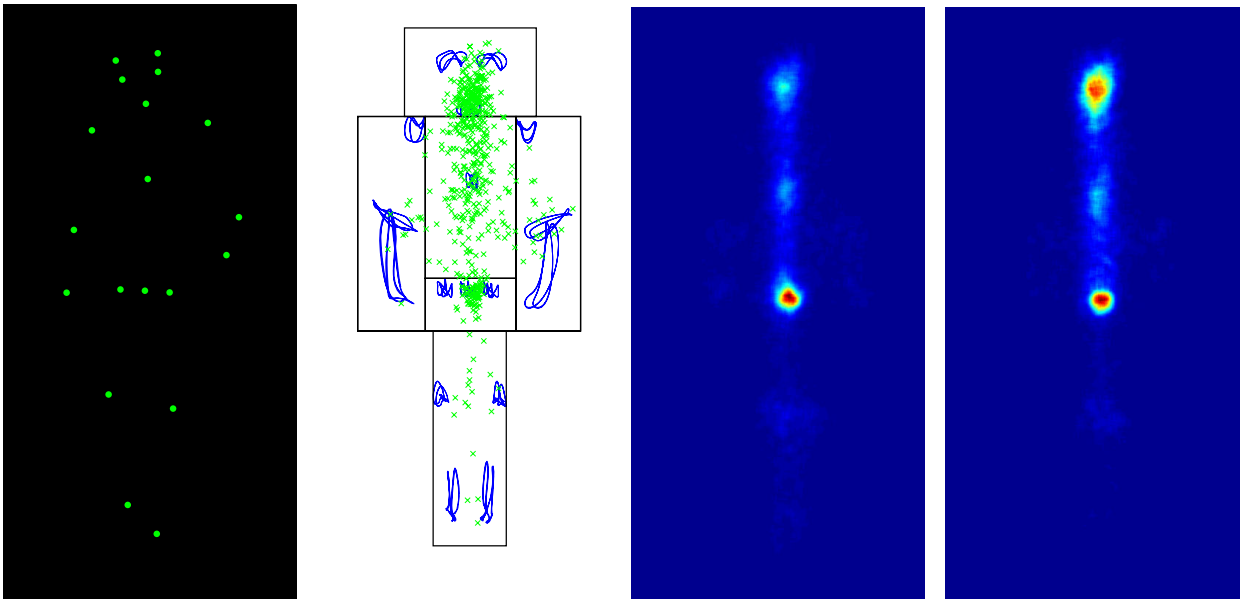
The results of this preliminary survey have motivated to further analyze the visual gaze behavior during an emotion recognition task. The gaze behavior is divided in saccades and fixations. Saccades are rapid, ballistic changes in eye position that occur at a rate of 3-4 per seconds depending on the task. During a saccade, the eye is blind. Visual information is acquired during the relatively long fixations between saccades. Location of the saccades are either bottom-up or top-down controlled. Although a partial interdependence exists between eye movement and spatial attention, attention is free to move independent of the location of the saccades, but eye movements require visual attention to focus on a target [63, 157].

Evidence exists in psychology that a strong interaction exists between the orienting of spatial attention and stimulus emotionality [148]. Contradictory results on this relation can be found in literature where the results depend on the demand on the attentional resource pool during the experiment. Despite that it is well known that emotional content of stimuli presented to the sensory system are principal indicators of the importance of these stimuli; therefore emotional stimuli, especially fear-related stimuli, pop out of an array of non-emotional stimuli. Furthermore, emotionality of a stimulus seems to be sensed before perceptual or categorical encoding of the stimuli themselves, which supports an early encoding of emotional stimuli. Hence, it appears that the mechanism for registration of the emotionality of a stimulus is fast, does possibly not require conscious awareness, and presumably poses minimal demand for attentional resources [148].

In the following experiment, the visual gaze behavior during an emotion recognition task is recorded with an eye tracker. Aim of the experiment is to find out which body parts are of special interest during judging the emotional state of a walker and if the gaze behavior differs between barely watching a walker and emotion recognition.

Experiment III: Analysis of the Visual Gaze Behavior with Eyetracking

The explorative analysis of the visual gaze behavior during watching emotive gait patterns is based on the data of the following experiment. A number of 12 participants watched PLWs expressing the emotions happy, angry, sad, and neutral. The PLW videos were created based on the recorded joint centers and the head markers of four walkers from the Munich Database. Each video lasted six seconds (resolution of 1024x768 pixels, 30Hz). The gaze was recorded with an eye tracker (1000 Hz) in collaboration with E. Wiese and J.



(a) (b) (c) (d)

Fig. 3.5: The number of fixations vary depending on the body segment and attention selection.
a) Light points display the joint centers in PLW videos.
b) Fixations (green crosses) mainly concentrate on the head, thorax, and hip region. The model for the body segments is derived from the marker positions over time (blue lines).
c) For bottom-up attention selection, mostly the region around the hip is fixated.
d) Preferred area of fixation shifts to the head for top-down attention selection during an emotion recognition task.

Zwickel of the Neuro-Cognitive Psychology Unit, Department Psychology, LMU Munich. Preliminary results in the supervised bachelor thesis [186] indicate that the upper body is more often observed than the lower body during an emotion recognition task.

In the first part of the experiment, participants watched a randomized sequence of the videos (4 walkers, 4 emotions, 5 repetitions) without any further instructions. In this case, attention selection is mainly guided by distinct stimuli based on primary visual features, so-called bottom-up. All participants recognized a human in the PLW. A relation between different walking styles and emotional states was not observed by the participants. In the second part of the experiment, the participants were asked to rate the expressed emotion of the PLWs. Response options were 'sad', 'angry', 'neutral', 'happy', and 'do not know'. In this case, the attention selection is top-down guided to search for task-relevant information. The overall recognition rate is 42% and separately for each emotion 50%, 34%, 51%, and 50% for angry, happy, neutral and sad, respectively. No assignment was chosen in 4% of the cases. The decrease in the recognition rate in comparison to previous results, see Tab. 3.4 and 3.5, is explained by poor representation of the head posture in the PLW. Although the tracked markers of the head were visualized, a forward bend of the head was hard to retrieve from the front view of the PLWs. This influences especially the expression of sadness, which is characterized by a forward bend of the head.

Calculation of the average fixation time during watching either a happy, neutral, angry,

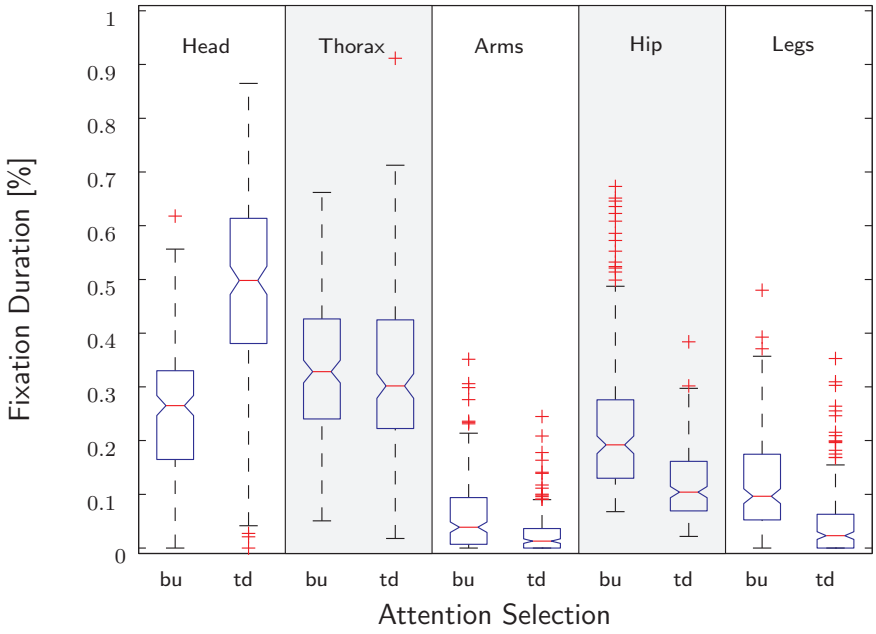
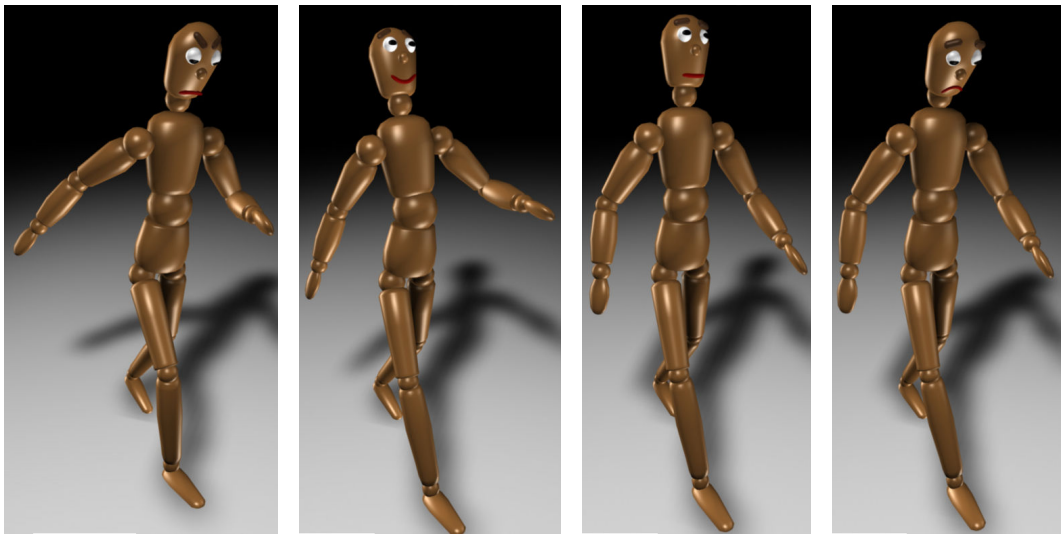


Fig. 3.6: Average fixation duration during top-down attention selection (td), here an emotion recognition task, increases for the region of the head in comparison to bottom-up attention selection (bu). The hip and the legs are less observed in the latter case.

or sad walker does not reveal any conspicuous differences for both cases bottom-up and top-down attention selection [186]. To analyze the spatial location of the fixations, the number of fixations for each body segment head, arms, thorax, hip and legs are calculated. Regions for each body segment are retrieved from the marker positions over time, see Fig 3.5 (b). Average number of fixations per video is 12.6 ± 5.6. The number of fixations is higher for the head, thorax and hip than for the leg and arm regions. During bottom-up attention selection, the hip is predominantly fixated. If the task switches to emotion recognition, preferred region of fixation shifts to the head even though facial expressions are not visible. Fig. 3.5 (c) and (d) illustrate the fixations of all participants for all walkers, emotions, and repetitions. Pixels of the videos which were seldom or not fixated are marked blue and pixels with a high number of fixations are marked red in the corresponding heat map.

Measured durations of fixations range between 0.168s and 2.975s with a mean of 0.529 ± 0.236s. Recording of one trial is included in the following analysis if it contains at least 5 fixations which leads to an upper limit of 1.2s for fixation duration. Durations of each region are added up for the five repetitions of watching a single PLW video and the percentage how long each region is fixated is jointly calculated. Fig. 3.6 shows the box-whisker plot for the fixation durations in percentage separately for bottom-up and top-down attention selection. Fixation duration is highest for the thorax, followed by the head and the hip for bottom-up attention selection. During the task of emotion recognition, the head is predominantly longer observed than the other body parts. In this case, the legs and the hip are less observed than during bottom-up attention selection. Note, Fig. 3.5 illustrates the absolute number of fixations per pixel and does not consider the size of each region. In contrast, the calculation of the fixation durations summarizes over all fixations within a region. Hence, although the region of the thorax is only marked in slight blue in Fig.



(a) (b) (c) (d)

Fig. 3.7: Authenticity is higher, the emotion is faster recognized, and the recognized emotion is more intense if facial expression and the gait express the same emotional state (a) anger, (b) happiness, (c) neutral, and (d) sadness [185].

3.5, summarizing over all fixations within this region leads to a higher value of fixation duration than for fixation duration of the hip. Following, results of Fig. 3.6 depend on the definition of the body segments. Still, it shows that the shift of attention from the center of the body to the head during emotion recognition is grounded in a lower observation of the lower body including the legs.

In accordance with the conducted survey, this experiment confirms that the head is predominantly observed during an emotion recognition task even if no facial expression is visible. This is controlled by a top-down attention selection mechanism.

3.3.3 The Interrelation between Expressive Gait Patterns and Facial Expressions on Emotion Perception

Usually, it is assumed that facial expressions are best to express emotions. For this reason, most scientific works concentrate on analyzing facial expressions in terms of basic emotional categories. However, if it comes to believability and authenticity, judgment relies not only on facial expressions but also on the congruence with the expression of further modalities.

Within this context, the question raises how emotion recognition relies on expressive gait patterns if the face expresses emotions simultaneously. Do different gait patterns influence the recognition at all? Which affective dimensions are preferable influenced by the walking style? Does authenticity decrease if facial expression and expressive gait patterns mismatch? To study all these questions, the following experiment has been conducted and the observers ratings are analyzed with regard to expressiveness, authenticity, and response time in the following.

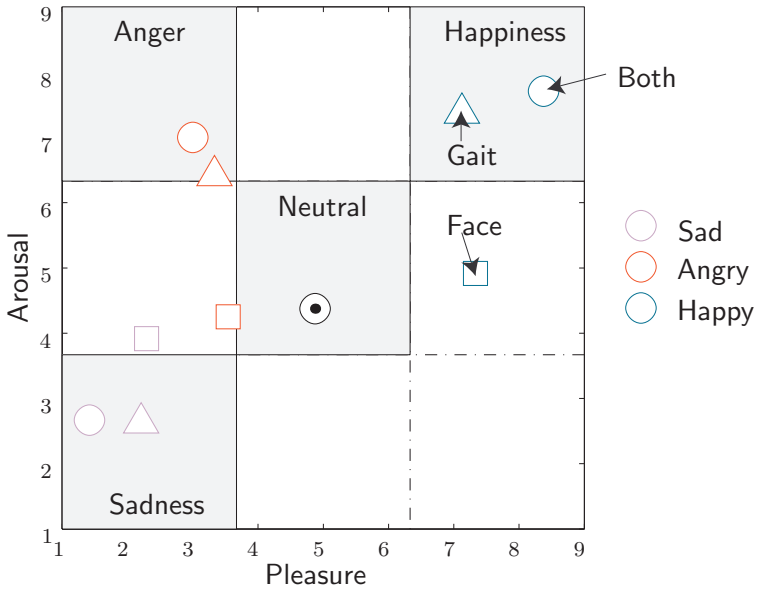


Fig. 3.8: If the animated puppet expresses the same emotion in gait and face (◯), average ratings of the participants lie in the expected regions. Additionally, the marker ⊙ indicates neutral expression of gait and face. Expressiveness is lower if emotions are only 1) expressed on the face combined with neutral walking (□) or 2) in the walking style combined with a neutral face (△).

Experiment IV: The Interaction of Expressive Walking and Facial Expressions

A number of 24 participants (5 female, 19 male, age: 27.2±3.4) took part in the experiment. Facial expressions for the animated puppet were designed as part of a supervised internship [185]. These expressions are anger, sadness, no emotion, and happiness, see Fig. 3.7. Four gait patterns expressing the same emotions were mapped to the animated puppet. The gait patterns were taken from the Munich database choosing the walker who expresses emotions best. Each facial expression was combined with one of the walking styles resulting in 16 videos lasting 30s (resolution of 640 × 480 pixels). The videos were shown in a randomized order to each participant. Participants rated simultaneously pleasure, activity and authenticity expressed by the animated puppet on a 9-item Likert scale. Furthermore, the response time was measured.

First, the design of the facial expressions is evaluated analyzing the response data for the videos showing the neutral gait pattern in combination with the four facial expressions. Different levels of pleasure are well expressed in the comic-style face. However, ratings for arousal differ only slightly for the expressions of the comic-style face. Fig. 3.8 shows the expected area for each expression and the average ratings. For videos showing different facial expressions in combination with a neutral gait pattern, average ratings for pleasure lie in the expected regions for pleasure. Differences in arousal are not sufficient to reach the expected regions for arousal. Thus, the static, comic-style facial expressions well communicate different levels of pleasure, but only slightly express different levels of arousal.

Subsequently, different walking styles in combination with the neutral face are examined.

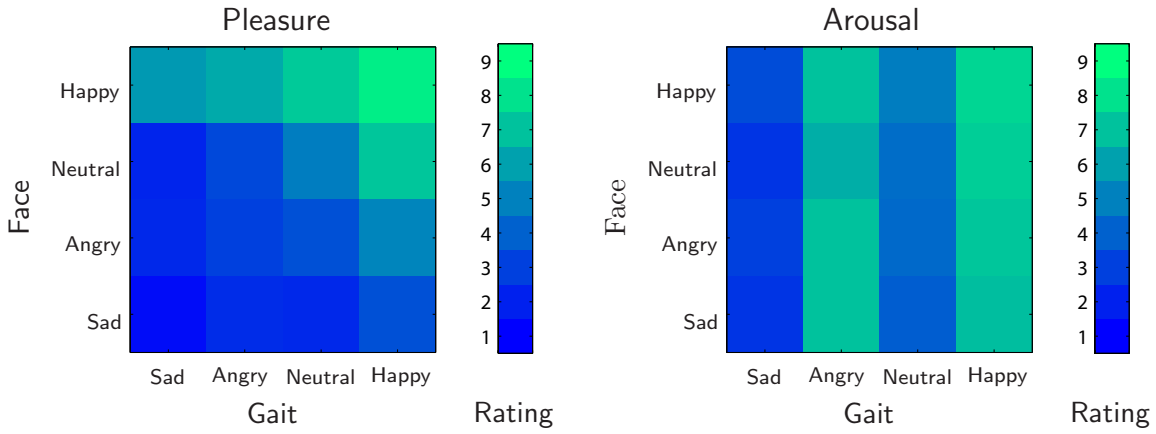


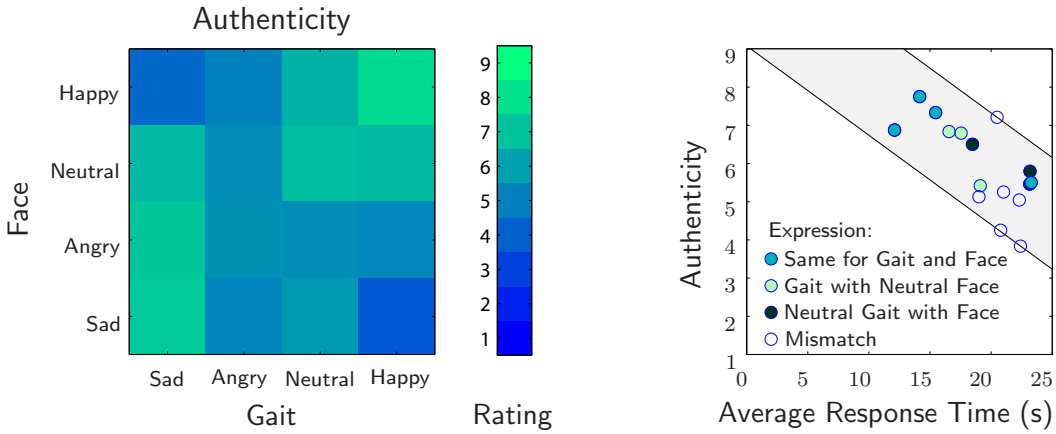
Fig. 3.9: Discordance leads to lower expressiveness for each emotion. As different levels of pleasure are similar recognizable for both modalities, the plot for pleasure follows a diagonal structure. However, the walking style communicates dominantly arousal so that changing the facial expression leads only to minor changes in arousal. This leads to a striated plot where highest average ratings were given to the fast walking styles which are anger and happiness (rating: 9 highest, 1 lowest).

Average ratings are plotted in Fig. 3.8. Expected regions are reached for both axes, pleasure and arousal. Recognition of the intended expression is further enhanced if facial expression and emotive gait pattern match, see Fig. 3.8.

If the expression of the face and the walking style mismatch, expression of pleasure and arousal decreases, see Fig. 3.9. Average ratings of the participants are closer to neutral expression for the case of incongruence. Furthermore, as the static comic-style facial expressions communicate only slightly differences in arousal, recognition of arousal is dominated by the walking style. Fast walking during happiness or in rage leads to higher recognized arousal.

For each video, participants were asked to rate how authentic the expressed emotion of the animated puppet is. Fig. 3.10 (a) and (b) summarize the results. An expressed emotion is perceived to be more authentic if facial expression and walking style match, see Fig. 3.10 (a). Based on the data of the experiment, the embodiment of the emotion anger seems to be less authentic in comparison to the other emotions. A possible reasons is that the design of the facial expression for anger is misleading. Most participants associate rather sadness than anger with the designed facial expression. Thus, believability is highest if the angry face is combined with sad walking leaving a discrepancy for combinations with the other emotions.

Plotting the overall response time over the degree of authenticity shows that the more authentic an emotion is the faster is its recognition, see Fig. 3.10 (b). If an emotion is expressed only either in the face or in the gait, response time increases. However, highest response time is measured for the cases in which the walking style mismatches with the facial expression. An exception is the expression of anger. Average response time for the combination of angry gait and angry face is with $23s$ almost twice as for the other congruent combinations. Fitting the data points with a Least Square algorithm to a linear



(a) (b)

Fig. 3.10: Authenticity of emotional expressions is related to congruence in the modalities and to the time for assignment of an emotional state to the animated puppet.
(a) If the facial expression and the walking style of the animated puppet express the same emotion, authenticity is highest.
(b) Furthermore, average response time for assigning pleasure, arousal, and authenticity decreases if the same emotion is expressed in gait and face.

function, provides an upper and a lower bound for the ratings. The area inside the bounds is gray shaded in Fig. 3.10 (b) and shows a negative relation between authenticity and time for estimating the emotional state of an approaching walker.

Summarizing the results of this experiment up, different levels of arousal are hardly expressed in a static, comic-style face. Expression of arousal can be improved by adding the appropriate expressive walking style. However, expression of pleasure can be improved only slightly by adding an appropriate walking style. Authenticity is increased and recognition of the expressed emotion is faster if the expression of the face and the gait match. Incongruence leads to less believability and taking longer time for making a decision on the expressed emotion.

As these results depend on the design of the facial expressions and the walking styles, generalization is limited to the combination of static facial expressions with different walking styles. Enhancing the animated puppet by dynamic facial expressions better communicates different levels of arousal via the face so that for such a case it is assumed that results for the dimension arousal will be more similar to the results of the dimension pleasure.

3.4 Summary

This chapter analyzes the perception and expression of emotions in gait patterns. It studies how well emotions can be decoded from variations in walking styles. These results serve as reference for automatic recognition in chapter 5 and 6. Furthermore, the visual gaze is investigated during observing emotive gait patterns. During the task of emotion recognition, attention shifts to the head even though no facial expressions are visible. Lastly, the interaction of facial expressions and emotive gait patterns is studied. Discordance leads to lower expressiveness. In the following, the contributions are discussed in more detail and in view of the state of the art.

Related studies in literature show that humans recognize emotions in walking better than chance, see Tab. 3.1. Average recognition rate ranges between 67% and 88%. Two gait databases are analyzed within this chapter. Recorded gait trials are either displayed as a PLW or as an animated puppet for human evaluation. Average recognition ranges between 42% and 63% for four categorical emotions. Recognition performance highly depends on the recognizability of the head posture. Lowest recognition is observed if the forward bend of the head is hardly observable. If the walker is shown from the side on the display, recognition ranges between 59% and 63%, which lies in the range of reported performance in [32]. Higher recognition rates as reported in [131] are only achieved in the experiments for walkers who express the emotions very well during recording. Comparing the recognition performance for the gait recordings of the Munich database with gait trials of the Emotive Motion Library and with in literature reported performance shows that expressiveness of the walkers lie in a similar range even though the emotion elicitation procedure has been eased to allow repetitive recording of each emotion. Hence, both databases are suitable for applying pattern recognition algorithms in chapter 5 and 6.

Furthermore, the following conclusions on the perception of emotions are drawn. First, humans retrieve emotional categories better than the gender of a walker from PLW displays. However, performance is lower than for identification of a known walker. Often, the emotion happiness is misclassified as anger and vice versa. This leads to the assumption that differences in arousal are easier to retrieve from gait than differences in pleasure. Analyzing the recognition performance separately for each dimension pleasure, arousal, and dominance shows that differences in arousal and dominance are better expressed and recognized in walking than differences in pleasure. Hence, utilizing the PAD model for analyzing affect is not only suitable for body postures, see [79], but also for whole-body motions such as gait. Second, observing the visual gaze behavior during watching PLW videos shows that the hip region is fixated predominantly. However during an emotion recognition task, the region of the head is primarily fixated even if no facial expressions are visible. This shift is controlled by a top-down attention selection mechanism. Third, if static facial expressions are available, expressiveness is increased if the body and the face express the same emotional state. Incongruence leads to less authenticity, less expressiveness, and longer time for judging the emotional state of the walker. Considering that static body postures provide more the specificity of emotions and body movements provide more the quantity or intensity of emotions [47], the influence of the walking style especially on the perception of arousal is reasonable.

Within this chapter, the human performance to detect emotions in gait is analyzed which is used later on as a baseline for comparison with automatic recognition. Besides utilizing a set of categorical emotions, experiments of this chapter indicate that the PAD model is also suitable for gait. A benefit of the PAD model for technical applications is its dimensional concept. Furthermore, humans observe in particular the upper body during an emotion recognition task. Hence, it is plausible to design also a classifier which is based only upon the joint angles of the upper body. As expressive walking styles influence the perception of emotions even if other modalities such as facial expressions are available, it is concluded to design appropriate walking patterns for a robot to increase authenticity. Furthermore, in multi-modal emotion recognition systems, the concurrence of the walking style and other modalities can provide additional information on the authenticity and intensity of an observed emotion.

3.5 Limitations

Limitations of the conducted studies are that the gait has not been divided in its posture and its dynamics, that the recorded gait patterns do not contain spontaneous emotions, and that interpretation is related to the applied emotion model and emotion definition. Furthermore, the study on the visual gaze during an emotion recognition task has been the first in this direction and is thus only descriptive. In a continuative study, it could also be considered to investigate if the presentation of the walker as PLW, animated puppet, or even animated puppet with facial expressions influences the perception. Limitation of the study about the interaction of expressive walking styles and facial expressions is that the facial expressions are static and that only these two modalities have been considered.Hence, this study can be extended to further modalities and dynamic facial expressions. Finally, this work concentrates on the body movement walking. Yet, knowledge can be transferred to other body motions and similarities as well as dissimilarities can be worked out for emotion expression in different body motions. In comparison to other modalities such as facial expressions and speech, expression of emotions in body motions is less investigated in psychology.

In summary, even though expression of emotions in gait is individual, humans can recognize emotional states from differences in walking styles. Several aspects concerning affective computing have been studied about human performance in more detail. Possible future directions may be directed towards a detailed analysis 1) how postural features interact with the expressiveness of motions, focus may in particular lie on the posture of the head, 2) to which degree gait expresses short-term emotions in comparison to long-term moods, and 3) of further whole-body motions to develop a generalizable coding system for emotions in motions.

4 Comparison of Inferential and Predictive Statistics in the Context of Gait Analysis

The previous chapter 3 analyzes aspects on human perception of emotions from observing gait patterns. The following chapters 5 and 6 study techniques from machine learning to automatically detect emotion in gait patterns. This chapter leads over from statistical analysis of human motion to automatic classification by investigating how mathematical techniques from both research fields relate to each other.

Traditionally, inferential statistics is applied in gait analysis. Within the last decade, several approaches have been undertaken to introduce methods from machine learning to gait analysis [21, 22]. Thereby, the main application is in therapeutic support. During the collaboration between computer scientists and researchers in both medicine and biomechanics, questions have come up how methods from machine learning relate to inferential techniques which are commonly applied in gait analysis.

Statistical inference supports the researcher to draw conclusions whether an effect is caused by a factor. It can be divided in four schools which are frequentist, Bayesian, fiducial, and likelihood inference. Frequentist approaches such as t-test and analysis of variance (ANOVA) are often applied in gait analysis. Yet, they have several disadvantages which are overcome by Bayesian inference. The latter is preferred in mathematical statistics and has been introduced to applied statistics. In machine learning, the algorithms estimate the cause from observations. Hence, inferential and predictive statistics analyze data in opposed direction and application of inferential or predictive statistics depends on the research problem. Both methods can only be applied to a data set if researchers are interested in inference as well as prediction. This can be the case in interdisciplinary research fields such as gait analysis, affective computing, or bioinformatics. Considering the recognition of emotions in gait patterns, researchers in psychology apply statistical inference to analyze how the gait is affected by emotional states. They provide a psychological foundation to retrieve the emotional state from the observation of gait patterns, which is studied in machine learning. This concept can be extended to any pattern recognition task for which the relation between cause and observation is not obvious and is an issue for scientific investigation. Thus, predictive analysis builds on a valid relation between cause and observation which can be verified by statistical inference. This describes the qualitative relation between inferential and predictive statistics, see Fig. 4.1 for illustration.

This chapter analyzes whether quantitative information in the context of predictive statistics can be gained from the results of inferential statistics. In machine learning, one major goal is to increase the recognition rate by improving the algorithms. From this point of view, it is of interest whether classification rates can be estimated from reported values of inferential statistics and, in doing so, serve as a baseline reference for advanced machine learning techniques. This chapter does not intend to rediscuss the advantages and disadvantages of Bayesian versus frequentist analysis, instead it concentrates on the

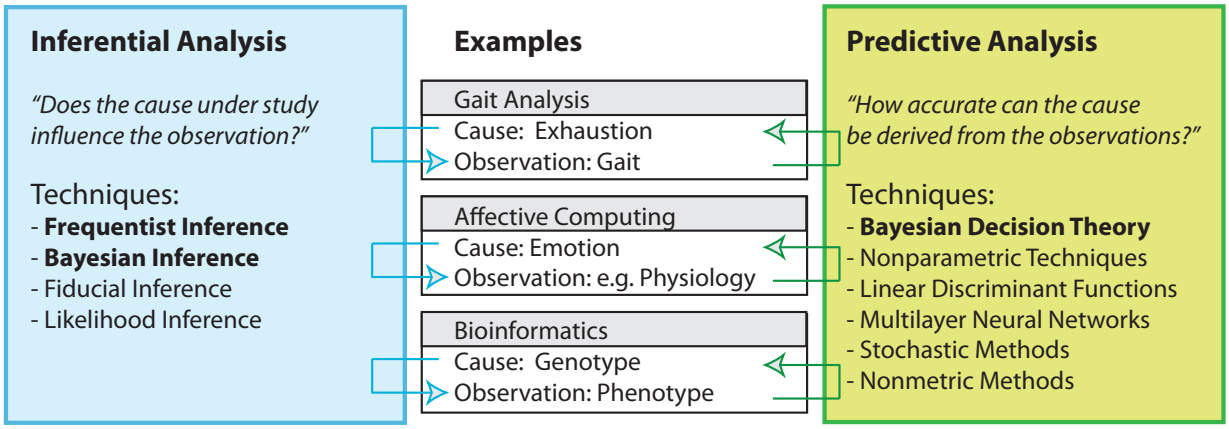


Fig. 4.1: Inferential statistics and predictive statistics analyze the relation between cause and observation in the opposed direction. The techniques compared within this chapter are marked bold.

interpretation of calculated, inferential statistics in the view of predictive statistics.

The investigation focuses on techniques which are univariate, assume normally distributed data, and divide the data in a limited number of groups. Thus, the t-test and ANOVA from inferential statistics are compared with a simple linear Bayes classifier from predictive statistics. The elaborated contributions within this chapter are 1) a mathematical relation between the statistics of an ANOVA and the classification rate of a linear Bayes classifier, which simplifies to a relation between the statistics of the t-test and the classification rate in the case of two classes, 2) upper and lower bounds for this relation, 3) a discussion of the consequence of this relation in using the considered statistics for feature selection, and 4) the application of the relation in the case of the dependent samples design.

The remainder of this chapter is organized as follows: First, section 4.1 defines the area of intersection which inferential as well as predictive analysis is applicable for. Then, the theoretical foundations of a selection of techniques from inferential and predictive statistics are reviewed. Statistics relevant for later comparison are introduced. A theoretical comparison is given in section 4.2 which shows that the statistics of the ANOVA relate to the accuracy of a linear Bayes classifier. Simplification of this relation for two classes leads to the relation between statistics of the t-test and the classification rate of a linear Bayes classifier. The following section 4.3 illustrates the relation for two and more classes. Thereby, upper and lower bounds for the relation between effect sizes of the ANOVA and the classification rate are derived. It also discusses its application for feature selection in machine learning and briefly goes into Bayesian inference. The application of the results for dependent samples test is derived in section 4.4 and the approximation of classification rates from reported test statistics is exemplified on a selected case study taken from gait analysis. Section 4.5 summarizes the investigated relation between selected techniques from inferential and predictive statistics. Finally, limitations of the comparison are discussed in section 4.6.

4.1 Selected Methods from Inferential and Predictive Statistics

Fig. 4.1 illustrates the relation between inferential and predictive statistics. A number of techniques exist for each analysis. The subdivision for predictive analysis is taken from the outline of [43]. The most common techniques, which are Bayesian and frequentist statistics, are selected from the techniques for inferential statistics for the following comparison. Regarding predictive analysis, Bayesian decision theory has the obvious advantage to be also based on the Bayesian theorem. Furthermore, it provides the optimal Bayes error rate $P_E^\star$ for known distributions. A graphical comparison of Bayesian decision theory and methods from frequentist analysis suggests a mathematical relation between these methods, see also Fig. 4.2 and 4.3 which consider Gaussian distributed data.

In the following, first, the area of intersection is defined in which predictive as well as inferential analysis may be applied within the research scope. It concludes with a list of terms which are used interchangeably. Then, frequentist and Bayesian inference are summarized. Focus lies on techniques which are univariate and which assume the data to be normally distributed. Afterwards, Bayesian decision theory is briefly introduced and the decision functions relevant for the comparison are given. Finally, a mathematical basis is described which motivates later comparison.

4.1.1 Area of Intersection and Terms

If researchers from social science, psychology, biology or medicine work together with engineers or computer scientists, the former is familiar with inferential analysis and the latter usually applies predictive analysis to the data. Examples for these collaborations are biomedical engineering, biomechanics, bioengineering, and affective computing. For illustration, three examples are described in more detail:

- Emotions are not only expressed in facial expressions and speech, but also in physiological parameters and body motions [32, 121, 125, 177, 178]. Affective computing aims to predict the emotional state of humans. As this research area requires intense collaboration between psychologists and computer scientists, the question comes up whether the results of a statistical analysis conducted by a psychologist can provide quantitative input to machine learning.

- Biomechanics in sports studies human motions during athletic activities. Bodily parameters under investigation are usually analyzed with statistical tools. Within this context, it is of interest whether methods from machine learning can provide additional information. Thus, several methods from machine learning have been applied [21, 22, 175]. A mathematical comparison of the two approaches which goes beyond application to a specific database has not yet been elaborated.

- Classification for computer aided diagnostics in e.g. mass spectrometry based proteomics or analysis of genomic data faces the challenge of a numerous number of features comparing to a limit number of samples [66, 94, 138, 164]. Thus, efficient

feature extraction is required. This approach is subdivided into filter and wrapper techniques. The former evaluates the quality of a feature based on its intrinsic characteristic, e.g. t-score, whereas the latter one estimates the accuracy of a specific classifier, e.g. Naive Bayes, applied to a feature subset. Still, no approximation exists how and which filter techniques relate to classification accuracy.

The area of intersection between predictive and inferential analysis is defined by

1. Predictive and inferential research interests on the data.
2. A limited number of outcome and independent variables.
3. Observations measured by an interval or ratio scale.
4. Groups subdivided by a nominal scale.

Within the above defined area of intersection, the focus of this investigation lies on clarification, in particular, on how the above mentioned methods relate to each other. This chapter analyzes univariate techniques which assume the data to be normally distributed. Thus, the conditions 3) and 4) relate to this chapter. They can be released by utilizing other techniques in inferential statistics, such as regression analysis or multivariate techniques.

Huberty and Olejnik state that application of predictive and inferential analysis cannot be used simultaneously [65]. This holds for the cases in which the criteria for the area of intersection are not fulfilled, e.g. evaluation of different drug treatments in medicine. However, due to increasing collaborations between researchers, new research fields emerge that require not only clarification regarding applied methods and algorithms but also regarding terms. Over time, different naming conventions have been evolved in each area of expertise. The following list briefly summarizes terms which are used interchangeably in literature:

- *Class*, *group*, level of an independent variable, condition.
- *Independent variable*, explanatory variable, exogenous variable, factor [1].
- *Feature*, observation, attribute, trait, characteristic, outcome variable[2], *dependent variable*, endogenous variable.
- Accuracy, *recognition rate*, success rate, *classification rate*.
- *Data set*, sample [3], sample set.
- Instance, *sample* [3], score.

Items in this chapter preferably used are marked italic.

[1]The term factor is used for an independent variable in a design which simultaneously evaluates the effect of two or more independent variables [143, p.738].

[2]Response variable associates outcome variable with grouping variable [65].

[3]Depending on the context, 'sample' refers either to a single entity of a population or to a subset of a population.

Test Decision	**State of the Nature**	
	H_0 **True**	H_A **True**
H_0 **rejected** (H_A accepted)	Type-I error α $p = P(\mathcal{X}\|H_0)$	Correct decision with power $(1-\beta)$
Fail to reject H_0	Correct decision (1-α)	Type-II error β $P(\mathcal{X}\|H_A)$

Tab. 4.1: Type of error which accompanies the test decision to accept or reject H_0.

4.1.2 Inferential Statistics

In contrast to descriptive statistics, statistical inference assumes randomized observations, utilizes the calculus of probabilities for modeling and description of stochastic processes, and aims to draw generally accepted conclusions about relations and processes underlying the data. Careful design of experiments is required so that inferential statistics can give valid conclusions. Frequentist inference and Bayesian inference are most commonly used in applied science. The former one was predominantly applied until the late 20th century. Since then, Bayesian inference is often preferred because it overcomes disadvantages of frequentist inference. The major difference between both schools is that frequentist inference consults the probability $P(\mathcal{X}|H)$ for decision making or in words how probable it is that the data $\mathcal{X}$ is observed if the hypothesis H is true. In contrast, Bayesian inference is based on the probability $P(H|\mathcal{X})$ which describes more the intention of the researcher to draw conclusions of the validity of a hypothesis H based on observations $\mathcal{X}$. Furthermore, Bayesian inference includes prior knowledge which can be either interpreted as a level of ignorance or information gained from prior studies.

The following summary considers univariate techniques, which assume the data to be normally distributed. The summary is based upon the textbooks [11, 27, 81, 136, 143].

Frequentist Inference

Frequentist inference analyzes statistically whether a hypothesis is true or not. In practice, a hypothesis cannot directly be validated. Thus, a Null hypothesis H_0 considered to be disproven is formulated. If H_0 is rejected with the Type-I error probability α, the alternative hypothesis H_A is accepted with the power $1-\beta$. Tab. 4.1 illustrates which type of error is made depending on the test decision. The Type-I error probability α describes the largest risk a researcher is willing to take of rejecting a true H_0. The Type-II error β refers to the conditional probability that the test decision is to retain H_0 although H_A is true.

Following Fisher's view of null hypothesis significance testing (NHST), an upper limit for the Type-I error probability α is defined before the experiment [49, 50]. Common values are 5%, 1% and 0.1%. During the statistical test, a test statistic is computed based on the data set. The estimated Type I error probability $p = P(\mathcal{X}|H_0)$[4] is derived. Then,

[4] $P(\mathcal{X}|H_i)$ refers to the conditional probability that the data set $\mathcal{X}$ is taken from a population Ξ for which the hypothesis H_i is valid.

the p-value is compared with its upper limit α. If it does not exceed its upper limit, H_0 is rejected and the alternate hypothesis H_A is accepted. Despite its easy application, the three major drawbacks of this approach are the following: 1) H_0 can only be disproven but not proven, 2) an estimate of the size of the effect itself is missing, and 3) the relation between number of samples and p-value is ignored.

Power analysis was introduced by Neyman and Pearson and overcomes these limitations by introducing the parameters power and effect size δ [115, 120]. In this approach, both H_0 and H_A are two competing hypotheses. Power refers to the probability to detect H_A under the condition that H_A is true. Thus, the power equals $1 - \beta$. For power analysis, a specific number of samples n is required to guarantee a certain test power. The number of samples n is determined a priory by $n = f(\alpha, \beta, \delta)$, in which the effect size δ is estimated based on previous studies. Power analysis extends Fisher's view of NHST in the way that a fix number of samples is determined to detect an effect of a certain effect size and defines power and α values a priori.

A combined version of both approaches is to follow Fisher's view of NHST and to report additionally the observed effect sizes to support meta-analysis. Reporting effect sizes has been advocated in several articles in social and biological science [29, 48, 78, 114]. However, it is not recommended to report post hoc obtained power values [92].

NHST is still an often applied statistical approach in biology, psychology, biomechanics, and biomedical statistics. Reasons are that NHST is an over decades established technique in applied statistics, that it considers random aberration by the use of stochastic calculus, and that it supports the process of decision making with probabilistic theory. Characteristic for the frequentist approach is to assume known variables of the random process which generates the data, and to calculate the probability to observe the data $\mathcal{X}$ if this process is true. Criticisms of this approach are that 1) not the probability $P(H_i|\mathcal{X})$ instead $P(\mathcal{X}|H_i)$ is consulted for decision making, 2) statistical significance does not automatically imply practical significance, and that 3) a dichotomous decision making process is intended [29]. Due to these criticisms, correct and adequate use of NHST has been advocated in several articles [28, 78, 112], which has led to a decrease in misuses during the last decade [145].

T-Test The t-test for two independent data sets $\mathcal{X}_1$ and $\mathcal{X}_2$ evaluates the hypothesis if the two data sets originate from two populations with different mean values μ_1 and μ_2. Hence, the null hypothesis is $H_0 : \mu_1 = \mu_2 = \mu$ and the alternative hypothesis is $H_A : \mu_1 \neq \mu_2$ for an independent, two-sided test.

In frequentist inference, the probability $P(\mathcal{X}|H_0)$ is calculated that the data $\mathcal{X}$ or data with a larger $\hat{t}$ statistic is observed under the assumption that H_0 is true. The statistic $\hat{t}$ is Student's t-distributed. For this concept, the following assumptions are made about the two population distributions Ξ_1 and Ξ_2:

- Random and independent selection of each sample from the population it represents,
- Homogeneity of variance (HOV): $\sigma_1^2 = \sigma_2^2 = \sigma^2$,
- Gaussian distributions: $\Xi_i \sim \mathcal{N}(\mu_i, \sigma^2)$.

To calculate the probability $P(\mathcal{X}|H_0)$, the $\hat{t}$ statistic[5] is computed

$$\hat{t} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{s_e^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{with} \quad s_e^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \ , \tag{4.1}$$

where s_e^2 is the pooled variance estimate, s_1^2 and s_2^2 the estimated variances for each data set $\mathcal{X}_i$, $\bar{x}_i$ the sample mean, and n_i the sample size of group i. To obtain $P(\mathcal{X}|H_0)$, the probability $1 - P(t \leq \hat{t})$ needs to be calculated. Since the t-test goes back to the times in which access to computers was rare, traditionally $\hat{t}$ is looked up in tables and explicit calculation of $P(\mathcal{X}|H_0)$ is avoided. Thus, the hypothesis H_0 is rejected with error probability α, if

$$\hat{t} > t_{\nu,1-\alpha/2} \ ,$$

with the degree of freedom ν. If the variances are homogeneous, ν equals $n_1 + n_2 - 2$ otherwise $\nu = n_2 - 1$ for $n_2 > n_1$. The t-test is robust against the Type-I error α, but not against the power [136]. As the Type-II error reaches a minimum for equal sample sizes n_i, it is usually recommended that 1) $n_1 \approx n_2$, 2) $n_i \gtrapprox 25$, and 3) a two-sided test is conducted. If $n_1 = n_2 = n$, Eq. 4.1 simplifies to

$$\hat{t} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2 + s_2^2}{n}}} \quad \text{with} \quad s_e^2 = \frac{s_1^2 + s_2^2}{2} \ .$$

Explanations which lead to rejecting H_0 are:

- Improper collection of sample sets resulting in observing an effect even though no effect exists.
- Application of the t-test even though its assumptions are not fulfilled.
- Observation of the effect occurred by chance.
- The effect really exists.

If the research question requires to compare the means μ_i of more than two classes, repetitive application of the t-test increases the probability to make a Type-I error. Thus, it is recommended to apply an analysis of variance (ANOVA) in this case.

ANOVA The single-factor ANOVA[6] for c independent sample sets $\mathcal{X}_1$,...,$\mathcal{X}_c$ with $c \geq 2$ evaluates the hypothesis, if at least two of the sample sets originate from populations with different means. Hence, the null hypothesis is $H_0 : \mu_1 = \mu_2 = ... = \mu_c = \mu$ and the

[5] If the variance σ^2 would be known, the Gauss test could be used instead.

[6] The single-factor ANOVA for independent sample sets is also referred to as the single-factor between-subjects ANOVA, the completely randomized single-factor ANOVA, the simple ANOVA, and the one-way ANOVA. The independent sample sets design is also known as independent-groups, between-subjects, between-groups and randomized design [143].

alternative hypothesis H_A states that at least one of the sample sets originates from a population with a different mean.

Similar to the t-test, the ANOVA assumes that H_0 is true and calculates the probability $P(\mathcal{X}|H_0)$ that the sample sets $\mathcal{X}_1, ..., \mathcal{X}_c$ are drawn from one common population. The ANOVA is based on the same assumptions as the t-test:

- Random and independent selection of each sample from the population Ξ_i,
- Homogeneity of variances (HOV): $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_c^2 = \sigma^2$,
- Gaussian distributions: $\Xi_i \sim \mathcal{N}(\mu_i, \sigma^2)$.

The variability in the data is subdivided in between-groups variability and within-groups variability. The F-measure is the ratio of between-groups variance s_{BG}^2 to within-groups variance s_{WG}^2. If all sample sets $\mathcal{X}_1$,...,$\mathcal{X}_c$ originate from one common distribution, the variability within the sample sets should be approximately the same as the variability between the sample means $\bar{x}_i$ and, thus, $F \approx 1$.

The $\hat{F}$ statistic is estimated by

$$\hat{F} = \frac{MS_{BG}}{MS_{WG}} \ , \tag{4.2}$$

with the mean MS_{WG} of the within-groups sum of squares SS_{WG}:

$$s_{WG}^2 = MS_{WG} = \frac{1}{N-c} SS_{WG} = \frac{1}{N-c} \sum_{i,j} (x_{ij} - \bar{x}_i)^2 = \frac{1}{N-c} \sum_i s_i^2 (n_i - 1) \ , \tag{4.3}$$

and with the mean MS_{BG} of the between-groups sum of squares SS_{BG}:

$$s_{BG}^2 = MS_{BG} = \frac{1}{c-1} SS_{BG} = \frac{1}{c-1} \sum_i n_i (\bar{x}_i - \bar{x})^2 \ , \tag{4.4}$$

where $N = \sum_{i=1}^{c} n_i$ is the total number of samples, $\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$ is the mean of group i, and $\bar{x} = \frac{1}{N} \sum_{i=1}^{c} n_i \bar{x}_i$ is the mean of the total data set. Furthermore, the total sum of squares is defined as $SS_T = SS_{BG} + SS_{WG}$.

The $\hat{F}$ statistic follows the Fisher-distribution with $\nu_{BG} = c - 1$ and $\nu_{WG} = N - c$ degrees of freedom. Similar to the t-test, the probability $P(\mathcal{X}|H_0)$ is traditionally not calculated, instead tabled F values are compared. If

$$\hat{F} > F_{\nu_{BG}, \nu_{WG}, 1-\alpha}$$

the null hypothesis H_0 is rejected and it is concluded that at least one sample set originates from a distribution with a different mean. The ANOVA itself does neither indicate which sample set nor how many means differ. For this purpose, multiple comparisons are usually conducted afterwards.

If sample sizes $n_1 = n_2 = \cdots = n$ are equal, violations of HOV have little influence on the test outcome and the power reaches a maximum. Furthermore, if the samples of all groups originate from a single common distribution $N(\mu, \sigma_T)$, the total variance σ_T is the sum of the within-groups and between-groups variance $\sigma_T^2 = \sigma_{WG}^2 + \sigma_{BG}^2$.

Effect Size Effect sizes provide an estimate of the magnitude of an effect and are essential for meta-analysis, because they are standardized measures independent of N [29, 78, 114]. The two dimensionless classes of effect sizes are differentiated in standardized mean differences and correlation coefficients. The former is used for categorical levels of an independent variable, and the latter for a continuous independent variable.

Depending on the context, numerous effect sizes have been proposed [77, 114]. Most commonly the standardized mean differences Hedges' g and Cohen's d index are reported for t-test [27, 114]. The δ index of the population expresses the difference between the means of the two populations in units of their variability

$$\delta = \left| \frac{\mu_1 - \mu_2}{\sigma} \right| . \tag{4.5}$$

The estimated d index is

$$d = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_e^2}} \right| . \tag{4.6}$$

The d index, also required for power analysis, is only practically reasonable, if the assumptions of homogeneity of variance and Gaussian distribution hold. Furthermore, Cohen in [27] suggests ranges for small, medium, and large effects, see Tab. 4.2. The relation between the d index and the test statistic $\hat{t}$ is:

$$d = \hat{t}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} . \tag{4.7}$$

As noted in [114], the effect size correlation r

$$r = \frac{\hat{t}}{\sqrt{\hat{t} + df}} ,$$

which is usually calculated for each continuous independent variable, can by converted to Cohen's d index by

$$d = \frac{2r}{\sqrt{1 - r^2}} .$$

For the one-way ANOVA, the most commonly used measures of the magnitude of an effect are ω^2, η^2 and Cohen's f index [27, 30, 143]. For c populations with different means μ_i and common variance within the populations $\sigma^2 = \sigma_{WG}^2$, the f index of the population refers to as

$$f = \frac{\sigma_{BG}}{\sigma_{WG}} \quad \text{with} \quad \sigma_{BG}^2 = \frac{1}{c}\sum_{i=1}^{c}(\mu_i - \mu)^2 . \tag{4.8}$$

It serves as a generalization of the δ index for more than two classes. The estimated $\hat{f}$

	Range		
Measure	Small	Medium	Large
d	$.2 - .5$	$.5 - .8$	$> .8$
$\hat{\eta}^2$, $\hat{\omega}^2$	$.01 - .06$	$.06 - .14$	$> .14$
$\hat{f}$	$.1 - .25$	$.25 - .4$	$> .4$

Tab. 4.2: Range for small, medium, and large effects as suggested in [27].

index is calculated by

$$\hat{f} = \sqrt{\frac{\frac{c-1}{nc}(MS_{BG} - MS_{WG})}{MS_{WG}}} = \sqrt{\frac{c-1}{nc}(\hat{F} - 1)} \; . \tag{4.9}$$

If $\hat{f} = 0$, the group means $\bar{x}_i$ are equal. Value $\hat{f}$ increases as the ratio of the between-groups variability and within-groups variability gets larger. It should be further noted that the size of $\hat{f}$ is independent of the number of classes c and the number of samples n, which facilitates meta-analysis. This independence is achieved – in contrast to the $\hat{F}$ statistic – by multiplying with the normalization term $\frac{c-1}{nc}$. Furthermore, $\hat{f}$ is only valid if H_0 is rejected. Otherwise $\hat{f}^2$ can become negative if $\hat{F} < 1$. Summarizing, $\hat{f}^2$ considers the ratio of between-groups to within-groups variance.

The measure[7] ω^2 considers the ratio of between-groups variance to total variance:

$$\omega^2 = \frac{\sigma_{BG}}{\sigma_{BG} + \sigma_{WG}} \; . \tag{4.10}$$

Its estimation based on a limited sample set is

$$\hat{\omega}^2 = \frac{SS_{BG} - (c-1)MS_{WG}}{SS_T + MS_{WG}} \; .$$

Commonly, the values of $\hat{\omega}^2$ range between 0 and 1. However, $\hat{\omega}^2$ will be a negative number if $f < 1$. The relation between $\hat{f}^2$ and $\hat{\omega}^2$ is :

$$\hat{\omega}^2 = \frac{\hat{f}^2}{1 + \hat{f}^2} \; . \tag{4.11}$$

Another commonly used measure for the effect size is

$$\hat{\eta}^2 = \frac{SS_{BG}}{SS_T} \; , \tag{4.12}$$

although it is a more biased estimate than $\hat{\omega}^2$. For sufficient large N, $\hat{\eta}^2 \approx \hat{\omega}^2$. Tab. 4.2 lists ranges that are associated with the meaning of low, medium and large effect sizes in social science.

[7] In [27], Cohen employs the notation η^2 in Eq. 4.10 instead of ω^2.

Bayesian Inference

Frequentist inference has been widely used in applied statistics. It provides a concept for hypothesis testing that calculates probabilities $P(\mathcal{X}|H_i)$ how probable values of statistics could occur assuming the hypothesis H_i to be true. Disadvantages are that the according tests rely on assumptions about the data, prior knowledge can not be included, and that constraints of parameters are hard to integrate. These disadvantages are overcome by Bayesian inference that calculates the probability $P(H_i|\mathcal{X})$. This technique allows easier handling of distributions which are not Gaussian, includes prior knowledge in the framework, and can handle parameter constraints. These reasons lead to a preference of Bayesian inference compared to frequentist inference in applied statistics. Still, frequentist inference is the dominant approach in a number of research fields e.g. psychology, social science and biomechanics. The two techniques fiducial inference and likelihood inference are more related to Bayesian inference with non-informative priors than to frequentist inference. Yet, these methods are not as general as Bayesian inference [11, p.73].

In the following, Bayesian inference is briefly introduced. Then, analogies for the t-test and the ANOVA in the Bayesian framework are described.

Bayes Theorem for Inference Bayesian inference draws conclusions about the parameters $\boldsymbol{\Theta} = [\Theta_1, ..., \Theta_k]$ of the probability distribution which underlies a stochastic process. Frequentist inference assumes that these parameters $\boldsymbol{\Theta}$ of the stochastic process are known and fix. Under this assumption, it calculates the probability for observing $[x_1, ..., x_n] = \boldsymbol{x}$. In contrast, Bayesian inference assumes that the form of the distribution underlying the stochastic process is known but that the estimation of its parameters $\boldsymbol{\Theta}$ from observation $\boldsymbol{x}$ follows itself a stochastic process. The stochastic processes are often modeled as Gaussian distributions but the principle of Bayesian inference is not restricted to this distribution.

Given that $\boldsymbol{\Theta}$ has the prior distribution $p(\boldsymbol{\Theta})$ and that $\boldsymbol{x}$ has the distribution $p(\boldsymbol{x})$, the Bayes theorem calculates the conditional distribution

$$p(\boldsymbol{\Theta}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\boldsymbol{\Theta})p(\boldsymbol{\Theta})}{p(\boldsymbol{x})} \ , \tag{4.13}$$

where $p(\boldsymbol{\Theta}|\boldsymbol{x})$ is also called the posterior. If $p(\boldsymbol{x}|\boldsymbol{\Theta})$ is regarded as a function of $\boldsymbol{x}$, it is called the likelihood $l(\boldsymbol{\Theta}|\boldsymbol{x})$. Furthermore, $p(\boldsymbol{x})$ can be regarded as a normalization constant. Thus, the posterior is proportional to the product of the prior distribution and the likelihood

$$p(\boldsymbol{\Theta}|\boldsymbol{x}) \propto l(\boldsymbol{\Theta}|\boldsymbol{x})p(\boldsymbol{\Theta}) \ .$$

In comparison to the Bayesian classifier, described in section 4.1.3, both $p(\boldsymbol{\Theta}|\boldsymbol{x})$ and $l(\boldsymbol{\Theta}|\boldsymbol{x})$ are distributions and not probabilities.

As the likelihood modifies prior information, each sample set $\boldsymbol{x}$ updates the current knowledge by utilizing the Bayes theorem. This technique is suited for sequential updating the current knowledge on $\boldsymbol{\Theta}$. For an initial set of observations $\boldsymbol{x}^1$ and a new set of

observations $\boldsymbol{x}^2$, the posterior is proportional to

$$\begin{aligned} p(\boldsymbol{\Theta}|\boldsymbol{x}^2, \boldsymbol{x}^1) &\propto p(\boldsymbol{\Theta}) l(\boldsymbol{\Theta}|\boldsymbol{x}^1) l(\boldsymbol{\Theta}|\boldsymbol{x}^2) \\ &\propto p(\boldsymbol{\Theta}|\boldsymbol{x}^1) l(\boldsymbol{\Theta}|\boldsymbol{x}^2) \ . \end{aligned} \tag{4.14}$$

In the following, this concept is exemplified on drawing inferences about a single mean μ. The set of observations $\boldsymbol{x} = [x_1, ..., x_n]$ are sampled from a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$ with unknown mean μ and known variance σ^2. Bayesian inference aims to draw conclusions on μ based on $\boldsymbol{x}$. Therefore, the likelihood $l(\mu|\boldsymbol{x})$ is calculated which is $\mathcal{N}(\mu, \sigma^2/\sqrt{n})$ distributed for a Gaussian process with known variance

$$l(\mu|\boldsymbol{x}) \propto \exp\left[-\frac{n}{2\sigma^2}(\mu - \overline{x})\right] \ , \tag{4.15}$$

with $\overline{x}$ being the sample mean. If little is known a priory about μ, a non-informative prior is chosen for $p(\mu)$. Here, the non-informative prior is a locally uniform distribution and the relation 4.15 is not altered by the prior. Thus, the posterior distribution is

$$p(\mu|\boldsymbol{x}) = \frac{1}{\sqrt{2\pi\frac{\sigma^2}{n}}} \exp\left[-\frac{n}{2\sigma^2}(\mu - \overline{x})\right] \ .$$

To draw inferential conclusions, probabilities are calculated whether μ is smaller or bigger as a reference value or whether μ is within a defined interval. In comparison to a Bayesian classifier, the posterior models the distribution of the μ of the process and not the occurrence of samples.

In the following, analogies to the t-test and the ANOVA in Bayesian inference are summarized. As the concept of Bayesian inference is more flexible than the concept of frequentist inference, these analogies are not as frequently used in Bayesian inference than the t-test or ANOVA in frequentist inference. The intention to include these methods within this chapter is to elaborate similarities and dissimilarities between using Bayes theorem for inference or prediction and to work out whether the relations between results of the t-test and ANOVA with Bayesian classification rates are in relation to the analogies of t-test and ANOVA in Bayesian inference.

Bayesian Inference for the Difference of Two Means Assuming a Known Common Variance The analogy to the t-test in Bayesian inferences compares two means μ_1 and μ_2. The observations $\boldsymbol{x}^{(1)} = [x_1^{(1)}, ..., x_{n_1}^{(1)}]$ are drawn from a population which is $\mathcal{N}(\mu_1, \sigma^2)$ distributed and the observations $\boldsymbol{x}^{(2)} = [x_1^{(2)}, ..., x_{n_2}^{(2)}]$ from a population which is $\mathcal{N}(\mu_2, \sigma^2)$ distributed. The corresponding sample mean values are $\overline{x}_1 = n_1^{-1} \sum_{i=1}^{n_1} x_i^{(1)}$ and $\overline{x}_2 = n_2^{-1} \sum_{i=1}^{n_2} x_i^{(2)}$. The sample variance s^2 equals the pooled variance estimate in Eq. 4.1. The joint posterior distribution is given by

$$p(\mu_1, \mu_2, \sigma^2|\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) = p(\sigma^2|s^2) p(\mu_1|\sigma^2, \overline{x}_1) p(\mu_2|\sigma^2, \overline{x}_2) \ ,$$

if μ_1, μ_2, and $\log\sigma$ are independent and the priors a locally uniformly distributed. Considering the difference $\mu_2 - \mu_1$, the joint distribution is

$$p(\mu_2 - \mu_1, \sigma^2|\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) = p(\sigma^2|s^2)p(\mu_2 - \mu_1|\sigma^2, \overline{x}_2 - \overline{x}_1)$$

The final posterior density function for this case is

$$p(\mu_2 - \mu_1|\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) = \frac{[s^2(n_1^{-1} + n_2^{-1})]^{-1/2}}{B(\frac{1}{2}, \frac{1}{2}\nu)\sqrt{\nu}} \left\{1 + \frac{[(\mu_2 - \mu_1) - (\overline{x}_2 - \overline{y}_1)]^2}{\nu s^2(n_1^{-1} + n_2^{-1})}\right\}^{-\frac{1}{2}(\nu+1)} \tag{4.16}$$

with the degree of freedom $\nu = n_1 + n_2 - 2$ and the complete beta function $B(a, b) = \Gamma(a)\Gamma(b)/(\Gamma(a + b)$. The gamma function is given by $\Gamma(a) = \int_0^{\inf} t^{a-1}e^{-t}dt$ for $a > 0$ and $\Gamma(0.5) = \pi$. Comparing Eq. 4.16 with the density function of a random variable y which is t-distributed with mean μ, scaling κ and ν degrees of freedom

$$p(y|\mu, \kappa, \nu) = \frac{1}{B(\frac{\nu}{2}, \frac{1}{2})\sqrt{\kappa}} \left[1 + \frac{(y - \mu)^2}{\kappa}\right]^{-\frac{\nu+1}{2}} \quad \text{for} \quad y \sim t(\mu, \kappa, \nu)$$

shows that $\mu_2 - \mu_1$ is $t(\overline{x}_2 - \overline{x}_1, s^2(n_1^{-1} + n_2^{-1}), n_1 + n_2 - 2)$-distributed. To apply the standard t-distribution $t(0, 1, n_1 + n_2 - 2)$ with the density function

$$p(\hat{t}|\nu) = \frac{1}{B(\frac{\nu}{2}, \frac{1}{2})\sqrt{\nu}} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} ,$$

the according transformation is

$$\hat{t} = \frac{(\mu_2 - \mu_1) - (\overline{x}_2 - \overline{x}_1)}{[s^2(n_1^{-1} + n_2^{-1})]^{-1/2}} . \tag{4.17}$$

Assuming that the Null hypothesis $\mu_2 = \mu_1$ is true, the calculated $\hat{t}$ statistic coincide with the $\hat{t}$-statistic for the t-test in frequentist inference, see Eq. 4.1, but the interpretation differs. Here, inference is based on the probability that the true difference between μ_2 and μ_1 is zero observing $\overline{x}_1$ and $\overline{x}_2$. Frequentist inference would assume $\mu_2 = \mu_1$ to be true and draws conclusions based on the probability to observe data with sample means $\overline{x}_1$ and $\overline{x}_2$.

Bayesian Inference for the Difference of Multiple Means Assuming a Known Common Variance Concerning a number of $c \geq 2$ means to be compared, a number of n_i samples $x^{(i)}$ are drawn from each distribution $\mathcal{N}(\mu_i, \sigma^2)$ with a total number of samples N. The sample mean for each set of observations is $\overline{x}_i$ and the common variance estimated from the samples $s^2 = \frac{1}{N-c}\sum_{i,j}(x_j^{(i)} - \bar{x}_i)^2$ is the within-groups variance given in Eq. 4.3. The vector $\boldsymbol{\mu} = [\mu_1, ..., \mu_c]$ is a set of mean values. Based on the posterior distribution $p(\boldsymbol{\mu}|\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(c)})$ for non-informative priors

$$p(\boldsymbol{\mu}|\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(c)}) \propto \left[1 + \frac{\sum_{i=1}^{c} n_i(\mu_i - \overline{x}_i)^2}{\nu s^2}\right]^{-\frac{\nu+1}{2}} \quad \text{with} \quad \nu = N - c ,$$

probabilities for particular sets of mean values $\boldsymbol{\mu}$ can be calculated. In practice, the comparison of two means is often more in the focus of interest than inference on a specific set of mean values $\boldsymbol{\mu}$.

4.1.3 Bayes Theorem in Predictive Statistics

Pattern recognition plays a major part in artificial intelligence. Algorithms draw conclusions from observations and attain knowledge. A number of techniques exist to assign class labels, such as nonparametric techniques, discriminant functions, and Bayesian decision theory. The central limit theorem says that the sum of arbitrarily but identically distributed random variables approximately follows a normal distribution. Hence, this gives a justification for the often used approach in applied science to model the data with a Gaussian distribution. Considering further that the t-test and ANOVA in frequentist inference rely on normally distributed data, this suggests to select the Bayesian decision theory for the intended comparison of techniques from inferential and predictive statistics. In the following, Bayesian decision theory is briefly introduced. The presentation is based upon its description in the textbooks [43, 72].

Bayesian Decision Theory

Given an observation or feature x and a set of memberships or classes $\boldsymbol{C} = \{\mathcal{C}_1, ..., \mathcal{C}_c\}$, the posterior probability $P(\mathcal{C}_i|x)$ describes the probability that the sample x belongs to the class C_i. Applying the Bayes theorem results in

$$P(\mathcal{C}_i|x) = \frac{p(x|\mathcal{C}_i)P(\mathcal{C}_i)}{p(x)} \ , \tag{4.18}$$

where $P(\mathcal{C}_i)$ is the prior probability, $p(x|\mathcal{C}_i)$ the class-conditional probability density function, and $p(x)$ the probability density function of x, given by $p(x) = \sum_{i=1}^{c} p(x|\mathcal{C}_i)P(\mathcal{C}_i)$. The Bayes classifier minimizes the average error rate by selecting the class $\mathcal{C}_i$ that maximizes $P(\mathcal{C}_i|x)$. This concept is extended in applications by using more than one feature, including actions, and considering more general loss functions than the error probability. Furthermore, it often performs well even if its assumption of independent features is violated [42, 163]. The intended comparison with inferential statistics focuses on univariate techniques. Therefore, a simple form of the Bayes classifier is considered in the following.

For the univariate case, the discriminant function $g_i(x)$ of the class $\mathcal{C}_i$ is

$$g_i(x) = \log P(\mathcal{C}_i) - \frac{1}{2}\left[\frac{(x-\mu_i)^2}{\sigma_i^2} + \log \sigma_i^2\right] \ ,$$

where $\mathcal{N}(\mu_i, \sigma_i^2)$ models the $\mathcal{C}_i$-conditional probability density function. If the priors $P(\mathcal{C}_i)$ are equal $\forall i$, the discriminant function can be further simplified to

$$g_i(x) = -\frac{1}{2}\left[\frac{(x-\mu_i)^2}{\sigma_i^2} + 2\log \sigma_i\right]$$

resulting in a quadratic decision border between each set of two classes. In addition, if it

is assumed that $\sigma_1 = \sigma_2 = ... = \sigma_c = \sigma$, the discriminant function is

$$g_i(x) = -\frac{1}{2}\left[\frac{(x-\mu_i)}{\sigma}\right]^2 , \tag{4.19}$$

and the decision borders between the classes are linear. The quadratic Bayes classifier always performs equal or better than the linear Bayes classifier if the population parameters μ_i and σ_i are known exactly. However, in general more samples are required to properly design a quadratic classifier than a linear classifier for practical applications [51]. Thus, the following analysis conducted for the linear Bayes classifier also provides a lower bound for the classification rate achievable with a quadratic Bayes classifier.

The classification rate of all samples for a single class is further denoted as $R(\mathcal{C}_i)$ and the average classification rate as

$$\bar{R} = \frac{\sum_{i=1}^{c} R(\mathcal{C}_i)}{c} . \tag{4.20}$$

4.1.4 Basis for Comparison

Selected techniques from inferential and predictive statistics have briefly been summarized. The focus lies on methods in inferential statistics which are often applied for gait analysis in biomechanics, and for various applications in affective computing. In doing so, the comparison is limited to univariate approaches which assume a normally distributed data.

Null hypothesis significance testing, such as t-test and ANOVA, draw inferences by consulting $P(\mathcal{X}|H_i)$ and is termed as a frequentist approach. Reason therefore is that the probability $P(\mathcal{X}|H_i)$ describes how probable the data $\mathcal{X}$ is observed if H_i is true. The principle is illustrated in Fig. 4.2 for a two-sided t-test. The difference of the sample means $\overline{x}_2 - \overline{x}_1$ is $\mathrm{t}(\mu_2 - \mu_1, s_e^2(n_1^{-1} + n_2^{-1}), \nu)$-distributed. Under the assumption that H_0 is true, $\mu_2 - \mu_1 = 0$ and the transformed $\hat{t}$ statistic is $\mathrm{t}(0, 1, \nu)$-distributed. Thus, t-test and ANOVA draw inferences about mean values and not about single samples. In contrast, the Bayes classifier computes the probabilities $p(x|\mathcal{C}_i)$ of single samples x to belong either to class $\mathcal{C}_1$ or $\mathcal{C}_2$. This is shown in Fig. 4.3.

Both approaches have in common the estimation of probabilities either for a single sample or for the complete sample set assuming a given hypothesis or class membership. Under specific assumptions, the t-test and linear Bayes classifier relate to each other, which is derived in the following section.

In contrast, Bayesian inference consults $P(H_i|X)$ to draw inferences about parameters $\boldsymbol{\Theta}$. An analogy to the traditional t-test exists which calculates the same t-statistic but the interpretation differs. Fig. 4.4 illustrates the calculation of the probability that the true difference between two means $|\mu_2 - \mu_1|$ lies within a certain range given the observed data $X = \{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}\}$.

In conclusion, predictive statistics aims to predict class membership of single samples whereas inferential statistics draws conclusions about parameters of the stochastic process which generates the data. When inferential statistics confirm information about the distributions underlying the stochastic process, is it then possible to estimate the performance of a simple Bayes classifier which predicts class membership of randomly drawn samples

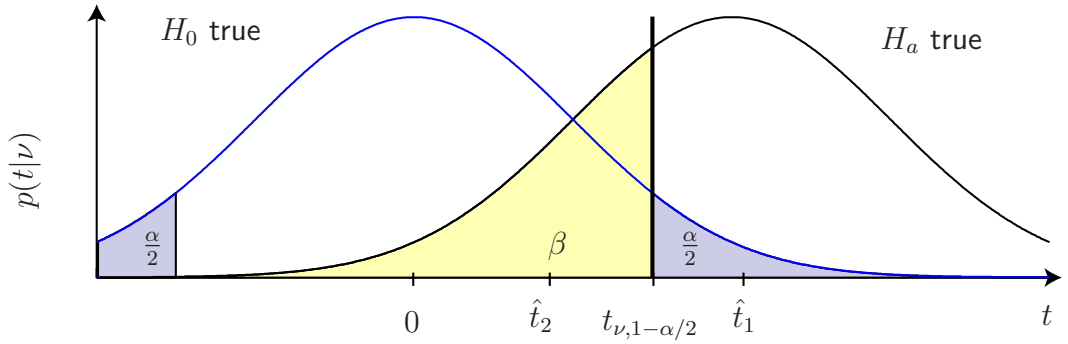


Fig. 4.2: The hypothesis H_0 is rejected with the test power $1-\beta$ and the Type-I error probability α if $\hat{t} > t_{\nu,1-\alpha/2}$ for a two-sided t-test, e.g. for $\hat{t}_1$ with $p(t|\nu) \sim t(0,1,\nu)$ for H_0. The test would fail to reject H_0 for e.g. $\hat{t}_2$.

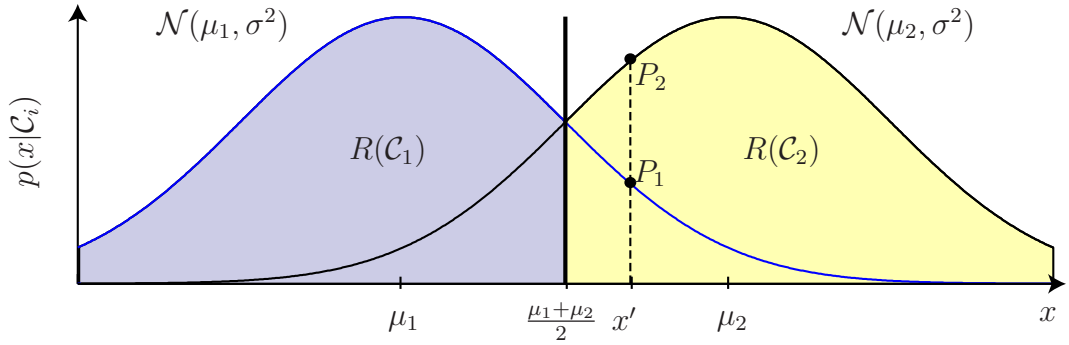


Fig. 4.3: The less the distributions $N(\mu_1,\sigma^2)$ for $\mathcal{C}_1$ and $N(\mu_2,\sigma^2)$ for $\mathcal{C}_2$ overlap, the higher the classification rate $\overline{R} = \frac{1}{2}[R(\mathcal{C}_1)+R(\mathcal{C}_2)]$ of a linear Bayes classifier is. The single sample x' belongs with probability P_1 to class $\mathcal{C}_1$ and with P_2 to class $\mathcal{C}_2$.

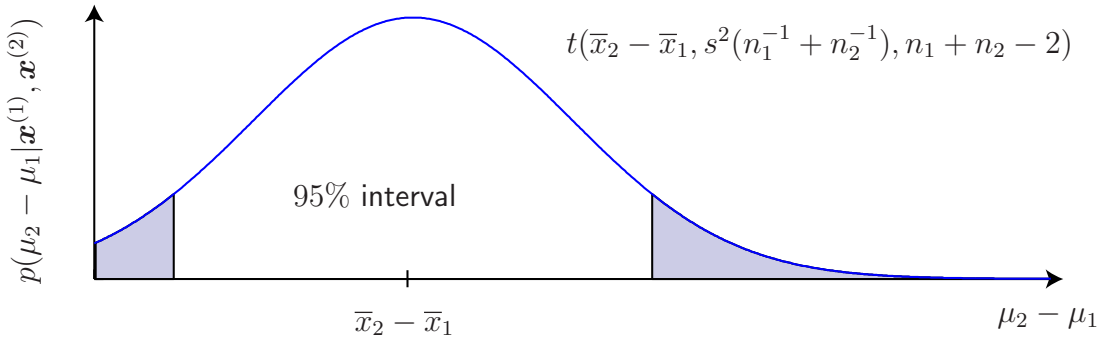


Fig. 4.4: Bayesian inference consults the probability that the true difference between two means $\mu_2-\mu_1$ lies e.g. within a 95% interval. If the prior is non-informative, the interval is centered around the difference of the sample means $\overline{x}_2-\overline{x}_1$.

from this stochastic process? This is investigated in the following sections.

Lastly, it should be further noted that drawing conclusions from classification rates to inferential statistics may be theoretically possible for a two-class problem. However, scientific inference requires a valid experimental setup to eliminate disturbing factors.

4.2 Mathematical Comparison

Cohen introduces the effect size measure d for the t-test in [27]. To visualize the magnitude of effects, he describes three measures $U_1 = \Phi(d)$, $U_2 = \Phi\left(\frac{d}{2}\right)$ and $U_3 = \frac{2\Phi(d/2)-1}{\Phi(d/2)}$ [27, p.23]. The measure U_2 explains the percentage of each distribution exceeding the other, see the blue and yellow shaded areas in Fig. 4.3. These areas match $R(\mathcal{C}_1)$ and $R(\mathcal{C}_2)$ for the Bayes classifier with a linear decision border. Thus, Cohen uses the classification rate $\bar{R}$ to illustrate the magnitude of effects. In the following, the assumptions for this relation and the mathematical deduction are summarized.

The combined assumptions for the t-test and a simplified Bayes classifier are

- Homogeneity of variance $\sigma_1 = \sigma_2 = \sigma$,
- Equal sample size $n_1 = n_2 = n$,
- Equal priors $P(\mathcal{C}_1) = P(\mathcal{C}_2) = P(\mathcal{C})$.

From these assumptions, $\bar{R} = R(\mathcal{C}_1) = R(\mathcal{C}_2)$ is deduced. For known population parameters μ_1, μ_2 with $\mu_2 > \mu_1$, and σ^2, the classification rate $\bar{R}(\Xi)$ is

$$\bar{R}(\Xi) = \frac{1}{2}\left[R(\mathcal{C}_1) + R(\mathcal{C}_2)\right] = \frac{1}{2}\left[\int_{-\infty}^{x^\star} p(x|\mathcal{C}_1) + \int_{x^\star}^{\infty} p(x|\mathcal{C}_2)\right] = \int_{-\infty}^{x^\star} p(x|\mathcal{C}_1) \ ,$$

with $x^\star = \frac{\mu_1+\mu_2}{2}$ being the solution for $g_1 = g_2$. Furthermore, $p(x|\mathcal{C}_1)$ is $\mathcal{N}(\mu_1|\sigma^2)$-distributed and $p(x|\mathcal{C}_2)$ is $\mathcal{N}(\mu_2|\sigma^2)$-distributed. With Φ being the standard normal distribution $\mathcal{N}(0,1)$, the following relation is derived

$$\bar{R}(\Xi) = \Phi\left(\frac{x^\star - \mu_1}{\sigma}\right) = \Phi\left(\frac{\mu_2 - \mu_1}{2\sigma}\right) \ .$$

Considering the definition of the effect size δ for populations, see Eq. 4.5, and that the absolute difference of the means is relevant $|\mu_2 - \mu_1|$, results in the following relation

$$\bar{R}(\Xi) = \Phi\left(\frac{|\mu_2 - \mu_1|}{2\sigma}\right) = \Phi\left(\frac{\delta}{2}\right) \ . \tag{4.21}$$

The according relation for sampled data is

$$\bar{R}(\mathcal{X}) = \Phi\left(\frac{d}{2}\right) = \Phi\left(\frac{\hat{t}}{\sqrt{2n}}\right) \ . \tag{4.22}$$

Thus, Cohen's measure U_2 exactly matches the classification rate $\bar{R}(\mathcal{X})$ of a Bayes classifier with a linear decision border. The relation between $\bar{R}(\mathcal{X})$ and d is also described in [180]. The relation between measured $\hat{t}$ values and $\bar{R}(\mathcal{X})$ depends additionally on n. This relation assumes that the parameters μ_1, μ_2, and σ are sufficiently well estimated from the data $\mathcal{X}$. In practice, the recognition rate of a Bayes classifier is often not calculated by estimating the distribution parameters followed by integrating, but by calculating the number of correctly classified samples. This summation and the reason that only a set of samples

from the population is observed leads to deviations from the theoretical value $\bar{R}(\Xi)$. For a sufficient large number of samples n, $\bar{R}(\mathcal{X})$ achieves $R(\Xi)$. Plots in the supervised diploma thesis [180] illustrate that calculating classification rates by either summation or integration differs only little.

In the following, this approach is extended for comparing more than two groups, $c > 2$. In this case, the average classification rate $\bar{R}$ of a multi-class problem is compared with the generalization of d, the f index. Without loss of generality, it is assumed that $\mu_1 \leq \mu_2 \leq .... \leq \mu_c$. Further assumptions for the derivation are:

- Homogeneity of variance $\sigma_1 = \sigma_2 = ...\sigma_c = \sigma$,
- Equal sample size $n_1 = n_2 = ... = n_c = n$,
- Equal priors $P(\mathcal{C}_1) = P(\mathcal{C}_2) = ...P(\mathcal{C}_c) = P(\mathcal{C})$.

The classification rate for each class $\bar{R}(\mathcal{C}_i)$ depends upon the dispersion of the means μ_i over their range between μ_1 and μ_c. Due to homogeneity of variances, the decision borders $\hat{x}_i$ are defined by

$$\hat{x}_i = \frac{\mu_{i+1} + \mu_i}{2} \quad \text{for} \quad i = 1...c-1 \ .$$

The classification rate $\bar{R}(\mathcal{C}_i)$ of a single class is given by

$$\bar{R}(\mathcal{C}_i) = \Phi\left(\frac{\mu_{i+1} - \mu_i}{2\sigma}\right) - \Phi\left(\frac{\mu_{i-1} - \mu_i}{2\sigma}\right) \quad \text{for} \quad i = 2...c-1 \ ,$$

where the classification rates $\bar{R}(\mathcal{C}_1)$ for the class with lowest mean and $\bar{R}(\mathcal{C}_c)$ for the class with largest mean are

$$\bar{R}(\mathcal{C}_1) = \Phi\left(\frac{\mu_2 - \mu_1}{2\sigma}\right) \quad \text{and} \quad \bar{R}(\mathcal{C}_c) = 1 - \Phi\left(\frac{\mu_{c-1} - \mu_c}{2\sigma}\right) \ .$$

Thus, for the average classification rate $\bar{R}(\Xi)$ follows

$$\bar{R}(\Xi) = \frac{1}{c}\left\{\Phi\left(\frac{\mu_2 - \mu_1}{2\sigma}\right) + \sum_{i=2}^{c-1}\left[\Phi\left(\frac{\mu_{i+1} - \mu_i}{2\sigma}\right) - \Phi\left(\frac{\mu_{i-1} - \mu_i}{2\sigma}\right)\right] + 1 - \Phi\left(\frac{\mu_{c-1} - \mu_c}{2\sigma}\right)\right\}$$

and written in terms of f by applying Eq. 4.8

$$\begin{aligned}\bar{R}(\Xi) &= \frac{1}{c}\left\{\Phi\left(\frac{\mu_2 - \mu_1}{2\sigma_{BG}}\ f\right) + \sum_{i=2}^{c-1}\left[\Phi\left(\frac{\mu_{i+1} - \mu_i}{2\sigma_{BG}}\ f\right) - \Phi\left(\frac{\mu_{i-1} - \mu_i}{2\sigma_{BG}}\ f\right)\right] + \right. \\ &+ \left. 1 - \Phi\left(\frac{\mu_{c-1} - \mu_c}{2\sigma_{BG}}\ f\right)\right\} \ . \end{aligned} \tag{4.23}$$

For the case of $c = 2$, Eq. 4.21 can deduced from Eq. 4.23 considering that f is a generalization of δ.

Characteristic for most real-world applications is that the standard deviations σ and σ_{BG} of the population are unknown and are estimated from samples. In pattern recognition, σ^2 is usually estimated by the pooled variance estimate s_p^2 for a linear Bayes classifier

$$s_p^2 = \frac{\sum_{i=1}^{c}((n_i-1)s_i^2)}{\sum_{i=1}^{c}(n_i-1)} = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2...(n_c-1)s_c^2}{n_1+n_2+...n_c-c} \ ,$$

which equals MS_{WG}, see Eq. 4.3. Thus, if MS_{WG} is reported, the classification rate of a linear Bayes classifier can easily be estimated by

$$\begin{aligned}\bar{R}(\mathcal{X}) &= \frac{1}{c}\Bigg\{\Phi\left(\frac{\bar{x}_2-\bar{x}_1}{2\sqrt{MS_{WG}}}\right)+ \\ &+ \sum_{i=2}^{c-1}\left[\Phi\left(\frac{\bar{x}_{i+1}-\bar{x}_i}{2\sqrt{MS_{WG}}}\right)-\Phi\left(\frac{\bar{x}_{i-1}-\bar{x}_i}{2\sqrt{MS_{WG}}}\right)\right]+ \\ &+ 1-\Phi\left(\frac{\bar{x}_{c-1}-\bar{x}_c}{2\sqrt{MS_{WG}}}\right)\Bigg\} \ . \end{aligned} \tag{4.24}$$

Furthermore, the relation between the $\hat{F}$ statistic and the classification rate of a linear Bayes classifier is given by

$$\begin{aligned}\bar{R}(\mathcal{X}) &= \frac{1}{c}\Bigg\{\Phi\left(\frac{\bar{x}_2-\bar{x}_1}{2\sqrt{MS_{BG}}}\ \sqrt{\hat{F}}\right)+ \\ &+ \sum_{i=2}^{c-1}\left[\Phi\left(\frac{\bar{x}_{i+1}-\bar{x}_i}{2\sqrt{MS_{BG}}}\ \sqrt{\hat{F}}\right)-\Phi\left(\frac{\bar{x}_{i-1}-\bar{x}_i}{2\sqrt{MS_{BG}}}\ \sqrt{\hat{F}}\right)\right]+ \\ &+ 1-\Phi\left(\frac{\bar{x}_{c-1}-\bar{x}_c}{2\sqrt{MS_{BG}}}\ \sqrt{\hat{F}}\right)\Bigg\} \ , \end{aligned} \tag{4.25}$$

and the relation between the effect size $\hat{f}$ and the classification rate by

$$\begin{aligned}\bar{R}(\mathcal{X}) &= \frac{1}{c}\Bigg\{\Phi\left(\frac{\bar{x}_2-\bar{x}_1}{2\sqrt{\frac{c-1}{N}(MS_{BG}-MS_{WG})}}\ \hat{f}\right)+ \\ &+ \sum_{i=2}^{c-1}\left[\Phi\left(\frac{\bar{x}_{i+1}-\bar{x}_i}{2\sqrt{\frac{c-1}{N}(MS_{BG}-MS_{WG})}}\ \hat{f}\right)-\Phi\left(\frac{\bar{x}_{i-1}-\bar{x}_i}{2\sqrt{\frac{c-1}{N}(MS_{BG}-MS_{WG})}}\ \hat{f}\right)\right]+ \\ &+ 1-\Phi\left(\frac{\bar{x}_{c-1}-\bar{x}_c}{2\sqrt{\frac{c-1}{N}(MS_{BG}-MS_{WG})}}\ \hat{f}\right)\Bigg\} \ . \end{aligned} \tag{4.26}$$

Transforming $\hat{f}$ to $\hat{\omega}$ would further establish the mathematical relation between $\hat{\omega}$ and the classification rate.

In summary, the classification rate of a linear Bayes classifier can be derived from reported test statistics of frequentist inference. For applying the derived Eqns. 4.21 – 4.26, it is assumed that the data sets fulfill the property of homogeneity of variance, that

Abbreviation	Term	Description
$\bar{R}(\Xi) = 1 - P_E^\star$	Classification rate	Average accuracy of the population Ξ calculated by integration using σ and μ_i
$\bar{R}(\mathcal{X})$	Classification rate	Average accuracy of a sample set $\mathcal{X}$ calculated by integration using s_i and m_i
$(\cdot)_q$		Accuracy of a quadratic Bayes classifier

Tab. 4.3: Symbols and naming convention for the average accuracy $\bar{R}$ of a Bayes classifier within this chapter.

the sample sizes of all groups are equal, that the data is Gaussian distributed, and that the parameters of the distributions are well estimated from the sample sets. On one hand, this provides an illustration how well several distributions are separable. On the other hand, this method facilitates the transfer of knowledge gained from inferential studies to machine learning. In doing so, the classification rate of a linear Bayes classifier can be estimated based on reported test statistics such as $\hat{t}$, d, and $\hat{f}$ using Eqns. 4.21 – 4.26. This procedure is especially advantageous if access to recorded databases from related studies is not public and an estimate how well methods from machine learning would perform for a certain problem is desirable. These estimations may serve as reference for more advanced techniques in machine learning.

Yet, the outcome of a hypothesis test is not qualitatively predictable from classification rates, because neither a power analysis is conducted nor the assumption of the t-test or ANOVA are tested in machine learning. Still, effect sizes can be estimated from classification rates, which can serve as parameter to estimate the appropriate number of samples n for a power analysis. Yet, common sense suggests that data sets which are well classifiable, should also differ qualitatively for the dependent variable in statistical inference.

4.3 Numerical Illustrations and Further Aspects

Mathematical relations between test statistics of frequentist inference and the classification rate of a linear Bayes classifier are derived above. These relations are illustrated in the following. The intention is to visualize the range for feasible classification rates given reported test statistics. The effect sizes δ and f are chosen as test statistics for t-test and ANOVA, respectively, because they are independent of the number of samples n and, therefore, facilitate plotting the relations. Furthermore, the relations are first illustrated for population parameters and then deviations are investigated for sample data. The approach to generate a database artificially, as in [180], is avoided to achieve independence from the quality how well sample distributions represent their populations. Thus, the following plots for the classification rate $\bar{R}(\Xi)$ over the effect sizes δ, f, and ω^2 refer to population parameters if not mentioned otherwise.

In the following, $\bar{R}$ denotes the accuracy of a linear Bayes classifier, see Tab. 4.3. For known population parameters, $\bar{R}(\Xi)$ equals $1 - P_E^\star$ with $P_E^\star$ being the Bayes error rate. Furthermore, the linear and quadratic Bayes classifier are generative classifiers and estimate

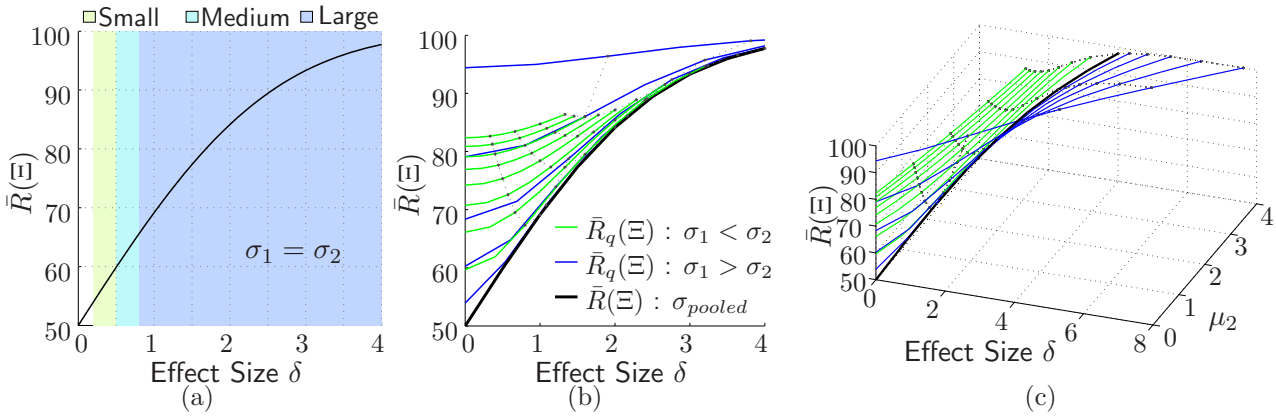


Fig. 4.5: (a) The classification rate $\bar{R}(\Xi)$ of a linear Bayes classifier increases with increasing effect size δ. (b) If the variances of the two populations differ, $\bar{R}(\Xi) = \Phi(\delta/2)$ still provides a lower bound for the recognition rate achievable with a quadratic Bayes classifier. (c) Additionally plotting μ_2 on a separate axis shows the plane spanned by different σ_1 and σ_2 for a quadratic Bayes classifier in more detail.

the population parameters σ_i and μ_i. As both estimates $\bar{x}_i$ and s_i are consistent, $\bar{R}(\mathcal{X})$ converges to $\bar{R}(\Xi)$ for increasing N. It should further be noted that recognition rates calculated in machine learning deviate from both $\bar{R}(\mathcal{X})$ and $\bar{R}(\Xi)$ in the way that leave-one-out or cross-fold validation is applied to the data set and, in doing so, the recognition rate is not calculated by integration but by summation over correctly classified samples. This leads to well-known differences between classification rates and recognition rates. The following illustrations refer only to classification rates assuming that the sample sets are representative for their populations.

First, the relation between statistics from frequentist inference and a linear Bayes classifier is illustrated for two classes. The larger the number of samples is the more reliable the estimated effect size is and the better $\bar{R}(\mathcal{X})$ is approximated. Second, illustrations are provided for more than two classes. In this case, the relation is more complex and both upper and lower bound are derived. For all illustrations, the means are ordered after $\mu_1 < \mu_2 < ...\mu_c$.

4.3.1 Illustrations for Case I: Two Classes

The t-test is applied to compare the means of two classes in frequentist inference. The Eqns. 4.21 and 4.22 assume that both populations have the variance $\sigma = \sigma_1 = \sigma_2$ in common. Fig. 4.5 (a) illustrates the relation between the effect size δ and the classification rate $\bar{R}(\Xi)$. The effect size δ describes the difference of the means $\Delta\mu = |\mu_2 - \mu_1|$ as multiples of the common standard deviation σ. Thus, an effect size of e.g. $\delta = 1$ would mean that the difference of the means $\Delta\mu$ equals their common standard deviation σ. In this case, the classification rate of a linear Bayes classifier is 69%. Considering the ranges for small, medium, and large effect sizes, as described in [27], the feasible classification rates $\bar{R}(\Xi)$ range as follows:

- The classification rate $\bar{R}(\Xi)$ is between 54% and 60% for small effect sizes δ.

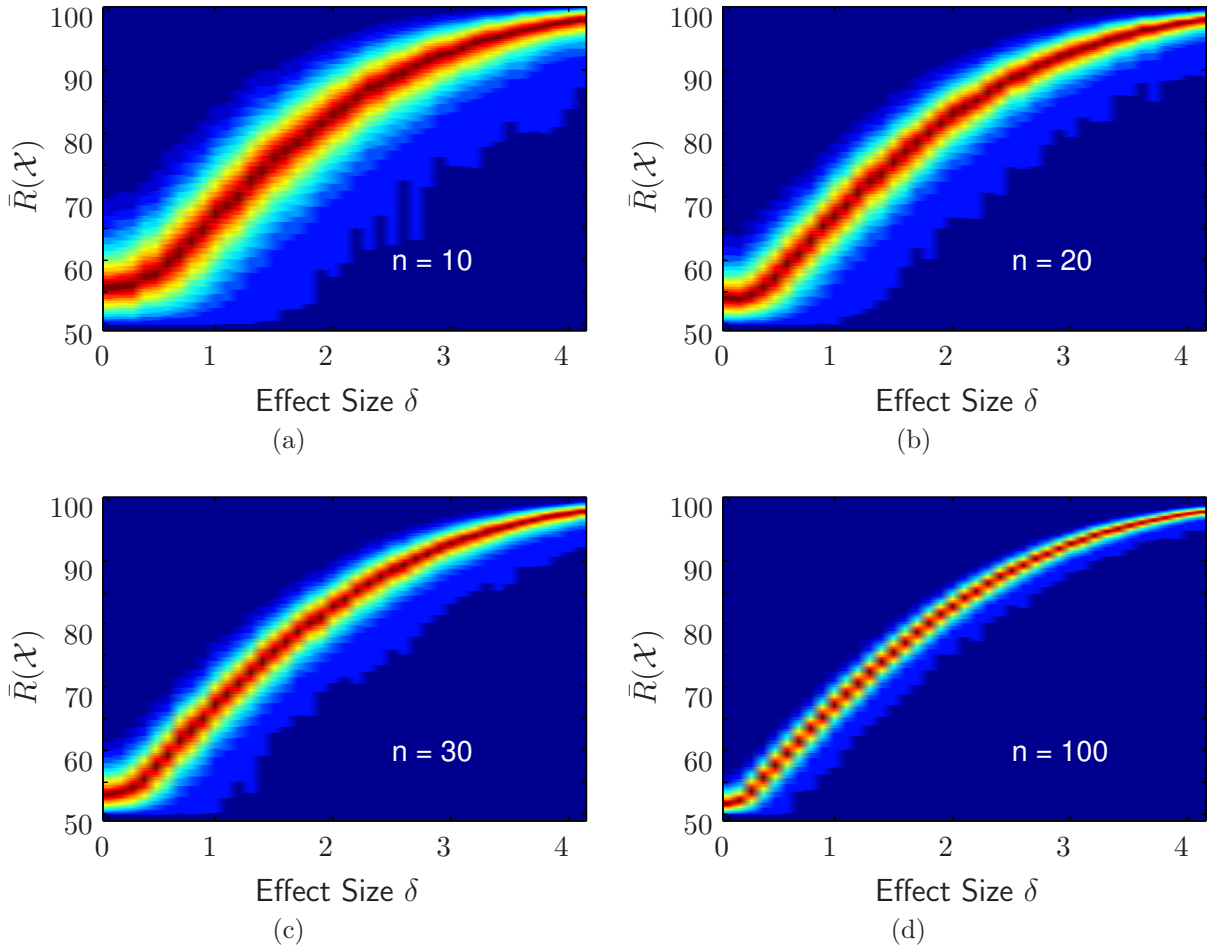


Fig. 4.6: A number of n samples is drawn from each of the two populations $\mathcal{N}(\mu_1, \sigma^2)$ and $\mathcal{N}(\mu_2, \sigma^2)$. Plotting the classification rate $\bar{R}(\mathcal{X})$ over the true effect size δ shows that the smaller the number of samples n is the larger deviates the measured effect size d and, hence, $\bar{R}(\mathcal{X}) = \Phi\left(\frac{d}{2}\right)$ from the true classification rate $\bar{R}(\Xi) = \Phi\left(\frac{\delta}{2}\right)$ of the two populations.

- The classification rate $\bar{R}(\Xi)$ is between 60% and 66% for medium effect sizes δ.
- The classification rate $\bar{R}(\Xi)$ is above 66% for large effect sizes δ.

This provides an estimate of $\bar{R}(\Xi)$ when the linear Bayes classifier is applied to single variable. In machine learning desirable classification rates above 90% may often not be feasible by considering only a single variable. Interactions between variables are not covered by a t-test and thus can not be estimated from statistics of the t-test.

If the standard deviation of the two populations σ_1 and σ_2 differ, the degree of freedom ν is decreased to $n - 1$ for the t-test. Fig. 4.5 (b) shows that Eq. 4.21 still provides an estimate of the recognition rate $\bar{R}(\Xi)$ for a linear Bayes classifier which uses the pooled variance estimate σ_{pooled}. Higher classification rates $\bar{R}_q(\Xi)$ are feasible for Bayes classifiers with quadratic decision borders. The more σ_1 and σ_2 differ the higher is the increase in the classification rate. This is illustrated in Fig. 4.5 (b) and (c). The

blue lines depict the cases for which $\sigma_1 < \sigma_2$. Illustrated examples are $\sigma_2 = h\sigma_1$ with $h \in \{0.05, 0.25, 0.45, 0.65, 0.85\}$. The green lines depict the cases for which $\sigma_1 > \sigma_2$. Illustrated cases are $\sigma_2 = h\sigma_1$ with $h \in \{1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}$. As for equal variances, the classification rate $\bar{R}(\Xi)$ increases with increasing difference $\Delta\mu$. The gray dotted lines depict the cases for $\Delta\mu \in \{\mu_1, 2\mu_1, 3\mu_1, 4\mu_1\}$ with $\Delta\mu = \mu_1$ starting on the left. It is concluded that in the case of different variances, $\Phi(\delta/2)$ still provides an estimate for $\bar{R}(\Xi)$ and a lower estimate for $\bar{R}_q(\Xi)$ [180].

If the parameters σ and μ_i are unknown, s and $\bar{x}_i$ are estimated from the samples. The estimates deviate from the true populations values leading to differences between the estimated effect size d and δ. Plotting the classification rate $\bar{R}(\mathcal{X})$ over the true value δ shows that the larger the number of samples n is the less differs the approximation of $R(\mathcal{X}) = \Phi(d/2)$ from $\Phi(\delta/2)$. This is shown in Fig. 4.6(a) - (d) with a varying number of samples n. For this plot, n samples are drawn for each population $\mathcal{N}(\mu_i, \sigma)$ 1500 times. The calculated classification rate $\bar{R}(\mathcal{X})$ is divided in 30 percentiles. Percentiles with high numbers of calculated $\bar{R}(\mathcal{X})$ from the samples are marked red. The higher the number of samples n is the less deviates the measured effect size d and, hence, $\bar{R}(\mathcal{X}) = \Phi\left(\frac{d}{2}\right)$ from the true classification rate $\bar{R}(\Xi) = \Phi\left(\frac{\delta}{2}\right)$ of the two populations.

4.3.2 Illustrations for Case II: More Than Two Classes

The ANOVA tests the null hypotheses whether the means of c classes are equal. If the null hypothesis is rejected, at least one mean differs from the others. The classification rate $\bar{R}$ of a linear Bayes classifier can be computed from reported statistics by applying Eqns. 4.23 - 4.26 for the distributions. Fig. 4.7 (a) illustrates the relation between the effect size f and $\bar{R}(\Xi)$. This plot is based on known population parameters. The larger the difference $\Delta = \mu_3 - \mu_1$, the higher is the effect size f and the classification rate $\bar{R}(\Xi)$ of a linear Bayes classifier. Furthermore, both values depend on the distributions of the means μ_1, μ_2, and μ_3 over their range $\Delta\mu$. Grey shaded lines illustrate this dependence for equal ranges $\Delta\mu = h\sigma$ with $h \in \{1, 2, 3, 4, 5\}$. Only value μ_2 varies on each of the gray shaded lines. Thus, the following summary can be given for small, medium, and large effect sizes f in the case of $c = 3$:

- The classification rate $\bar{R}(\Xi)$ is between 36% and 41% for small effect sizes f.
- The classification rate $\bar{R}(\Xi)$ is between 40% and 46% for medium effect sizes f.
- The classification rate $\bar{R}(\Xi)$ is above 44% for large effect sizes f.

The effect size ω^2 is scaled in comparison with f. The relation between ω^2 and $\bar{R}(\Xi)$ is plotted in Fig. 4.7 (b). Even if a maximum effect size of $\omega^2 = 1$ is reached, the classification rate $\bar{R}(\Xi)$ varies between 66% and 100% depending on the value of μ_2. Plotting $\bar{R}(\Xi)$ over the effect size $\hat{\eta}^2$ would result in a similar plot. The reason therefor is that $\hat{\eta}^2$ is only a slightly more biased estimate than $\hat{\omega}^2$, and $\hat{\eta}^2 \approx \hat{\omega}^2$ for large n.

Concerning the spread of the c means of the distributions, Cohen describes three patterns of variability in [27], which are minimum, intermediate and maximum. For this purpose, the normalized maximum range of the means is given by $D_{max} = \frac{\mu_c - \mu_1}{\sigma}$. The

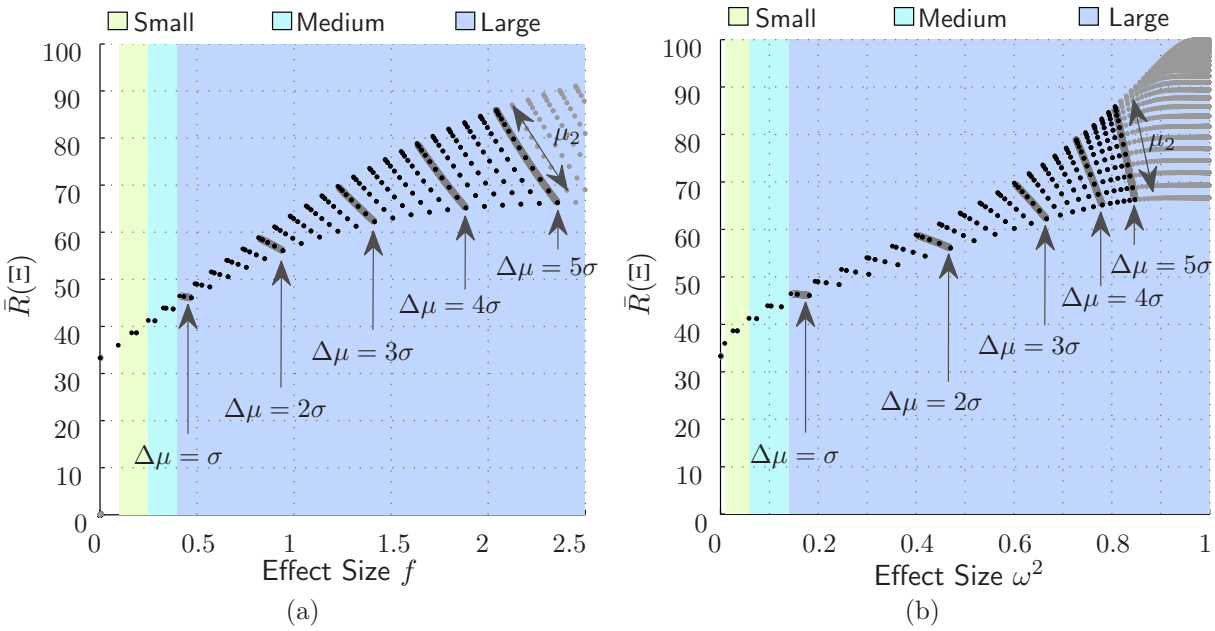


Fig. 4.7: (a) The classification rate $\bar{R}(\Xi)$ and the effect size f depend not only on the maximum range $\Delta = \mu_3 - \mu_1$ but also on the value of μ_1. (b) The effect size ω^2 is scaled in comparison to f and leads therefore to a different graphical form of the relation between effect size and classification rate. The graph for the effect size $\hat{\eta}^2$ is similar to this plot because $\hat{\eta}^2$ is only more biased than $\hat{\omega}^2$ and, for sufficient large n, $\hat{\omega}^2 \approx \hat{\eta}^2$.

remaining $c-2$ *means* are distributed over the maximum range D_{max}. The variability is categorized as follows, see Fig. 4.8:

- *Minimum variability:* The value of one mean is at each end point of D_{max}, the remaining $c-2$ means are at the midpoint. Then one has for the population parameters $f = D_{max}\sqrt{\frac{1}{2c}}$. The effect size f is minimal. The classification rate is maximal only for $c = 3$.

- *Intermediate variability:* The c means are equally spread over D_{max} and $f = \frac{D_{max}}{2}\sqrt{\frac{c+1}{3(c-1)}}$. For this case, the maximum of the classification rate is achieved.

- *Maximum variability:* The means are all at the end points of D_{max}. If the number of classes c is odd $f = D_{max}\frac{\sqrt{c^2-1}}{2c}$ and if c is even $f = \frac{1}{2}D_{max}$. In this case, f is maximal and the classification rate is minimal.

For the case of 3 classes, intermediate and minimum variability is the same. If $c > 3$, the intermediate variability defines an upper bound for the classification rate $\bar{R}(\Xi)$. Upper and lower bounds depending on the maximum range $\mu_c - \mu_1$ are derived in the following.

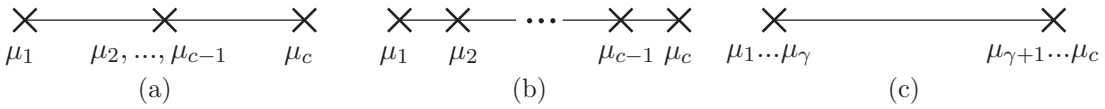


Fig. 4.8: In [27], three patterns of variability for c means are described: (a) minimum variability, (b) intermediate variability, and (c) maximum variability.

Given a maximum range of the means $\mu_c - \mu_1$, the intermediate variability provides an upper bound $\bar{R}_{UpperBound}(\Xi)$ for the classification rates. The effect size for this case is

$$f_{UpperBound} = \frac{D_{max}}{2}\sqrt{\frac{c+1}{3(c-1)}} = \frac{\mu_c - \mu_1}{2\sigma}\sqrt{\frac{c+1}{3(c-1)}}\ , \tag{4.27}$$

and the the corresponding maximum of $\bar{R}(\Xi)$ is

$$\bar{R}_{UpperBound}(\Xi) = \frac{1}{c}\left[2(c-1)\Phi\left(\frac{\mu_c - \mu_1}{2\sigma(c-1)}\right) - c + 2\right]\ . \tag{4.28}$$

The upper bound is illustrated as a red line for three, four, five, and six classes in Fig. 4.9. The red circle in the plots depicts the case $\mu_c - \mu_1 = 5\sigma$ for the upper bound. The yellow dots depict cases for which μ_2, ..., μ_{c-1} vary between the maximum range $D_{max} = 5$. This plane exemplifies that the larger the number of classes becomes the more complex the relation between the spread of the means and the values of f and $\bar{R}(\Xi)$ gets. Still, the planes have a lower bound which is given by the maximum variability.

Considering the case of maximum variability, a number of γ means lie at μ_1 and $c - \gamma$ means lie at μ_c with $\gamma \in \{1, ..., c-1\}$. For this case, the effect size $f_{LowerBound}$ of the population is given by

$$f_{LowerBound} = \frac{1}{\sigma\sqrt{c}}\sqrt{\gamma\left(\mu_1 - \mu_{LowerBound}\right)^2 + (c-\gamma)(\mu_c - \mu_{LowerBound})^2} \tag{4.29}$$

with

$$\mu_{LowerBound} = \frac{1}{c}[\gamma\mu_1 + (c-\gamma)\mu_c]\ , \tag{4.30}$$

The value $f_{LowerBound}$ is largest if the means are distributed between μ_1 and μ_c equally. Then, $f_{LowerBound} = \frac{D_{max}}{2}$ if c is even, and $f_{LowerBound} = D_{max}\frac{c^2-1}{2c}$ if c is odd. Independent of γ, the corresponding classification rate $\bar{R}_{LowerBound}(\Xi)$ is

$$\bar{R}_{LowerBound}(\Xi) = \frac{2}{c}\Phi\left(\frac{\mu_c - \mu_1}{2\sigma}\right)\ . \tag{4.31}$$

This rate $\bar{R}_{LowerBound}(\Xi)$ is the smallest value if D_{max} is hold fixed and the values of $c-2$ means are varied within the range between μ_1 and μ_c. The values of $\bar{R}_{LowerBound}(\Xi)$ are encircled in blue and black for $D_{max} = 5$ in Fig. 4.9. The black circle illustrates the case when the means are equally distributed on μ_1 and μ_c. In doing so, the effect size f is maximal [27]. The minimum classification rate $\bar{R}_{LowerBound}(\Xi)$ is independent of γ. Consequently, a number of $c/2$ lowest points with the coordinates $(f_{LowerBound}, \bar{R}_{LowerBound}(\Xi))$

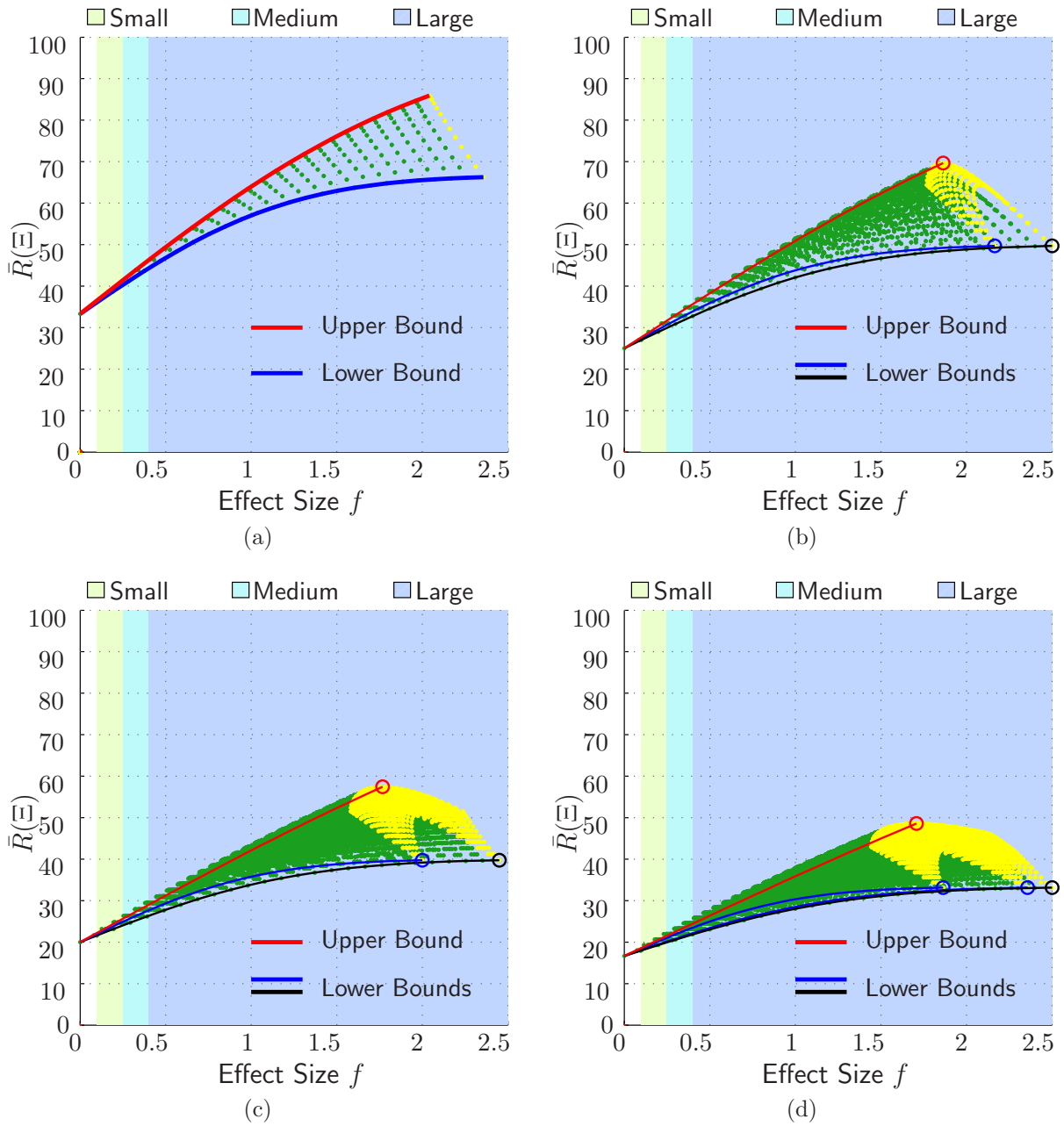


Fig. 4.9: The classification rate $\bar{R}(\Xi)$ depends not only on the effect size f but also on the distribution of the means between μ_1 and μ_c. The yellow points illustrate the plane on which the maximum range $D_{max} = \frac{\mu_c - \mu_1}{\sigma}$ is hold constant and the remaining $c-2$ means vary between μ_1 and μ_c. The number of classes are in (a) $c = 3$, in (b) $c = 4$, in (c) $c = 5$, and in (d) $c = 6$. The larger the number of classes becomes, the more complex is the plane. Furthermore, intermediate variability is related to maximum ability to classify and maximum variability to minimum ability to classify. Hence, possible classification rates for a given maximum range D_{max} are bounded by upper and lower bounds.

exist for each plane on which D_{max} is constant. Connecting these points leads to the lower bounds in Fig. 4.9. The black line denotes the case for which $c-2$ *means* are equally distributed on μ_1 and μ_c.

It should be noted that the lower and upper bounds define the minimum and maximum values for each plane upon which D_{max} is hold constant and the position of $c-2$ *means* is varied between μ_1 and μ_c. With increasing effect size f, the recognition rate $\bar{R}(\Xi)$ increases. Hence, the plane upon which D_{max} is constant shifts to higher F and $\bar{R}(\Xi)$ values. Planes with a larger value D_{max} overlap with the previous planes and points lie above the upper limit due to the form of the planes. Thus, the upper and lower bounds require D_{max} to be known. Only in the case of 3 classes, D_{max} can be unknown and still no points lie outside the upper bound. Reason therefor is that the plane upon which D_{max} is constant simplifies to a line for 3 classes.

Lastly, Fig. 4.9 provides an estimate on which classification rate $\bar{R}(\Xi)$ can be expected from small, medium, and large effect sizes f. Expectable classification rates $\bar{R}(\Xi)$ decrease with increasing number of classes for the small and medium effect sizes, e.g. $\bar{R}(\Xi)$ lies in the range of 40% and 46% for a medium effect size f for $c = 3$ but it lies within the range of 18% and 26% for $c = 6$.

However, the effect size of the population is usually unknown in applications. The estimated effect size $\hat{f}$ slightly underestimates the true value f. In the following, the upper and lower bounds are derived for estimated parameters, both $\bar{x}_i$ and s_i^2, and the difference between estimated bounds and the bounds for population parameters is illustrated in Fig. 4.10. Furthermore, the homogeneity of variances $s_1^2 = s_2^2 = ...s_c^2$ is assumed in the following.

The upper bound is given by

$$\hat{f}_{UpperBound} = \sqrt{\frac{(\bar{x}_c - \bar{x}_1)^2}{MS_{WG}} \cdot \frac{(c+1)}{12(c-1)} - \frac{c-1}{c \cdot n}} \tag{4.32}$$

and

$$\bar{R}_{UpperBound}(\mathcal{X}) = \frac{1}{c}\left[2(c-1)\Phi\left(\frac{\bar{x}_c - \bar{x}_1}{2(c-1)\sqrt{MS_{WG}}}\right) - c + 2\right] \ , \tag{4.33}$$

for estimated means $\bar{x}_i$ and common variance $s^2 = MS_{WG}$.

The estimated effect sizes $\hat{f}_{LowerBound}$ for the lower bound are

$$\hat{f}_{LowerBound} = \sqrt{\frac{\gamma(\bar{x}_1 - \bar{x}_{LowerBound})^2 + (c-\gamma)(\bar{x}_c - \bar{x}_{LowerBound})^2}{c \cdot MS_{WG}} - \frac{c-1}{c \cdot n}} \tag{4.34}$$

with

$$\bar{x}_{LowerBound} = \frac{1}{c}[\gamma\bar{x}_1 + (c-\gamma)\bar{x}_c] \ . \tag{4.35}$$

The corresponding classification rate $\bar{R}_{LowerBound}(\mathcal{X})$ is

$$\bar{R}_{LowerBound}(\mathcal{X}) = \frac{2}{c}\Phi\left(\frac{\bar{x}_c - \bar{x}_1}{2\sqrt{MS_{WG}}}\right) \ . \tag{4.36}$$

Fig. 4.10 illustrates the relation between $\bar{R}(\mathcal{X})$ and $\hat{f}$ for $c = 3$. The yellow dots exemplify the relation of $\bar{R}(\mathcal{X})$ and $\hat{f}$ based on n samples drawn from the distributions $\mathcal{N}(\mu_i, \sigma^2)$. For a small number of samples n, data points may lie slightly outside the bounds depending on how well the sample data is representative for the population, e.g. deviation from Gaussian distribution or homogeneity of variance. With increasing number of samples n, the bounds for the population match better the bounds for the estimates. Furthermore, the bounds for population or estimated parameters differ particularly in the range of small and medium effect sizes if n is small. The reason therefor is that the calculation of the classification rate $\bar{R}(\mathcal{X})$ does not distinguish between population parameters μ_i and σ and their estimates $\bar{x}_i$ and s, whereas the formulas for the effect size f differ. The value $\hat{f}$ is slightly lower than f for the same classification rate $\bar{R}(\mathcal{X})$ and for small values $\hat{f}$. The larger the effect and the larger the number of samples n is, the less differ the bounds for the estimates and the population parameters.

The interpretation in the context of estimating classification rates $\bar{R}(\mathcal{X})$ from reported $\hat{f}$ values is as follows. Exact calculation of $\bar{R}(\mathcal{X})$ can be achieved by means of the Eqns. 4.23 - 4.26. The plots 4.7 - 4.10 can be used to give an approximation of the classification rate $\bar{R}(\mathcal{X})$ without explicit calculation, e.g. in the cases if one is more interested in an estimate about the range of possible classification rates instead of exact values or if not all required values for Eq. 4.23 - 4.26 are reported. The plots are based upon the population parameters μ_i and σ. Hence, approximation of $\bar{R}(\mathcal{X})$ using these plots slightly underestimates the calculation of $\bar{R}(\mathcal{X})$ using the Eqns. 4.23 - 4.26 for small n and small $\hat{f}$.

Lastly, the derived equations for the upper and lower bounds can be converted to the effect size measures $\hat{\omega}^2$ using Eq. 4.11.

Summarizing this subsection, the relations between the effect size measure f and the classification rate $\bar{R}$ of a linear Bayes classifier have been investigated for more than two classes. For three classes, the classification rate $\bar{R}$ lies between 36% and 46% for small and medium effect sizes. The classification rate decreases if the number of classes increases for same f values. Figs. 4.7 - 4.10 illustrate this relation and can be used to estimate classification rates from reported f values graphically. Furthermore, upper and lower bounds are derived for the relation between $\bar{R}$ and f.

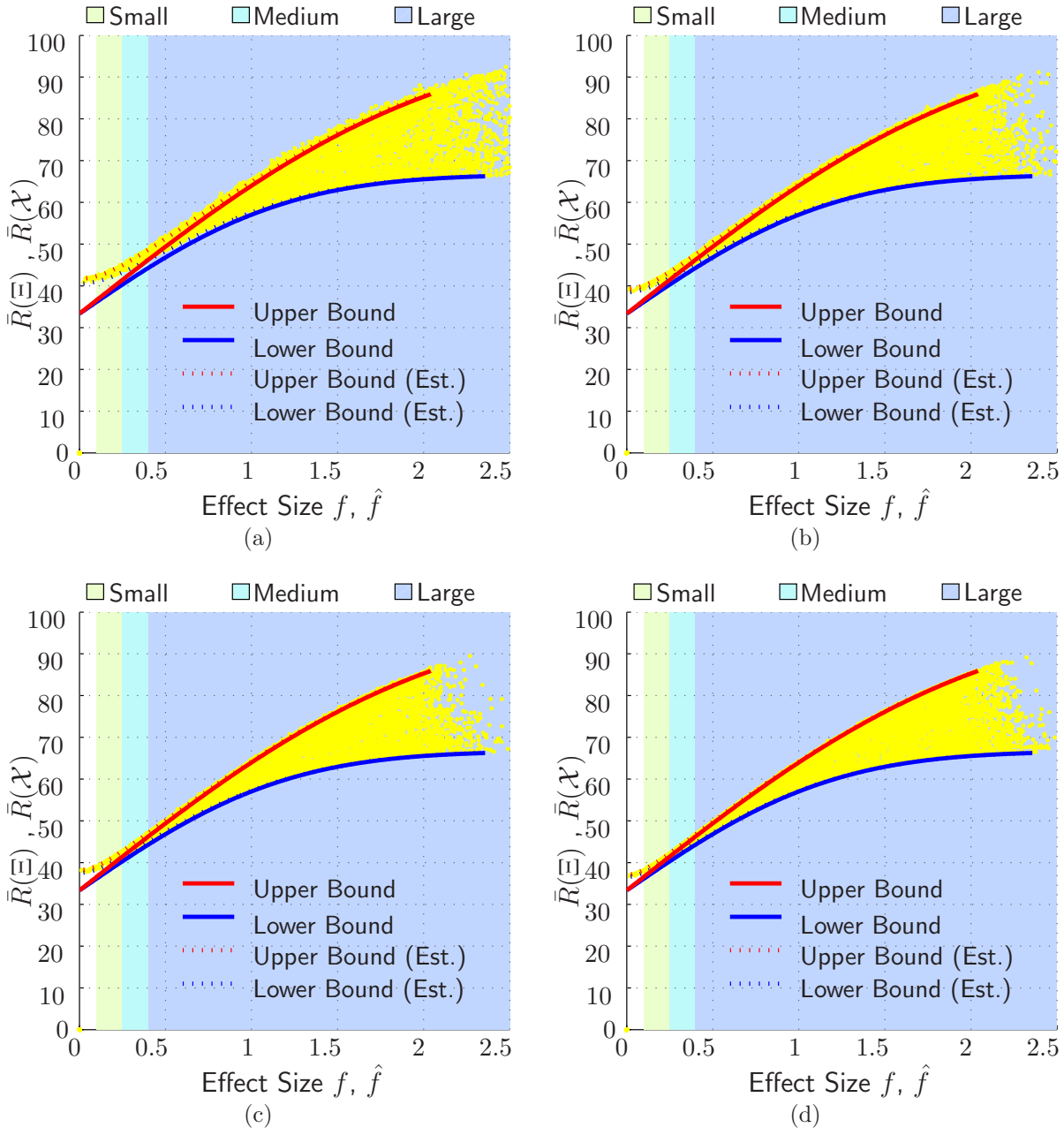


Fig. 4.10: The larger the number of samples n is the better the bounds of the population match the bounds of the estimates. Thus, for small and medium effect sizes and a small number of samples, $\bar{R}(\mathcal{X})$ would be larger than estimated from the bounds for the population. The relation between f and $\bar{R}(\mathcal{X})$ is illustrated for the cases (a) $n = 20$, (b) $n = 30$, (c) $n = 50$, and (d) $n = 100$.

4.3.3 Statistics for Feature Selection

Pattern recognition within machine learning aims to predict group memberships from observations. The process is generally subdivided in data extraction, dimension reduction, and classification [9, 43]. Dimension reduction is further subdivided in projection (e.g. PCA), compression (e.g. using information theory), and feature selection techniques. Saeys *et al.* in [138] give a comprehensive review on feature selection techniques. Feature selection is categorized in filter, wrapper, and embedded techniques. The former one is subdivided in univariate and multivariate approaches and also consults methods from NHST to select application relevant features. Univariate filter approaches, which ignore feature dependencies, are of special interest within the context of this chapter. In comparison to wrapper or embedded techniques, their advantages are easy use for high-dimensional data sets, fast computation, and independence from the classification algorithm.

T-test and ANOVA are widely used for feature selection although their statistical assumptions are not always verified or reported [66, 67, 94, 97, 138]. Either the p value for 2-class problems or the F statistic for multiple-class problems is calculated for each feature separately. Features with lowest p value or highest F statistic are selected for further classification. Several modifications of t-test exists which differ primarily in estimation of the pooled variance and non-parametric test have been introduced, which better deal with uncertainty about the true underlying distribution of the data [138]. Furthermore, the ANOVA is similar to linear discriminant analysis, because both approach are based on the ratio of between-variance to within-variance in the data.

According to Eq. 4.22, the t-value of the t-test relates nonlinearly to the Bayes error rate. This relation requires homogeneous variances, normally distributed data, and equal sample sizes. Under these conditions, the classification rate $\bar{R}$ of a linear Bayes classifier is directly proportional to t-values of the t-test. Thus, the lower the p-value of the t-test is for one feature, the higher is $\bar{R}$ for this feature. From this follows that if $p_{Feature1} < p_{Feature2}$ then $\bar{R}_{Feature1} > \bar{R}_{Feature2}$ is valid. In conclusion, results of the t-test used for feature selection are in relation to the Bayes error rate if the assumptions are valid.

The Bayes error rate and the $\hat{F}$ statistic of the ANOVA depend further on the distribution of the means, see Eq. 4.25. Therefore, the classification rate $\bar{R}_{Feature1}$ for a feature with a high $\hat{F}_{Feature1}$ value is not necessarily higher than the classification rate $\bar{R}_{Feature2}$ for a second feature with a lower $\hat{F}_{Feature2}$ value. This means graphically that the plot for the relation between a fix F or $\hat{F}$ value and the classification rate is not a point but at least a line or a plane. The form of the line or plane is defined by the additional dependence on the distribution of the means, see Figs. 4.7 - 4.10 for illustration of these planes. Still, upper and lower bounds exist given by the Eqns. 4.27 - 4.36. These bounds require additionally knowledge of the maximum range D_{max} of the means. Thus, possible classification rates for given F or $\hat{F}$ values are bounded. B. Guo and M. Nixon mention in their article [58] about gait feature subset selection by mutual information that a disadvantage of feature selection by a one-way ANOVA or correlation metrics is a lack of explicit or definite relation with classification accuracy or the Bayes error rate. The derived Eqns. 4.23 - 4.36 overcome this limitation which has accompanied using ANOVA as tool for feature selection.

The conclusions drawn for using F or t values to select features in machine learning can be transferred to further statistics for feature selection which can be converted to F or t

values or which are based upon these values.

4.3.4 Bayesian Inference

Bayesian inference intends to draw inferences about population parameters $\boldsymbol{\Theta}$ from the data $\mathcal{X}$. If means of two distributions with common variance σ^2 are compared, the same t-statistic is calculated as for traditional NHST. If this t-value is reported, the classification rate of a linear Bayes classifier can be calculated by applying Eq. 4.22. Usually the posterior probability $p(\mu_1, \mu_2, \sigma^2|\mathcal{X})$ would be calculated that the true difference of the means lies within a certain interval. In this way, the application of the Bayes theorem in inferential statistics differs from a Bayesian classifier which calculates the probability that a single sample is correctly classified.

It should be noted that further interesting relations exist between using the Bayes theorem in inferential statistic and machine learning, e.g. in the case of estimating a parameter, the bootstrap distribution equals a nonparametric, non-informative posterior distribution [61, p.272].

4.4 Relevance for Gait Analysis

Studies on gait in medicine and biomechanics often rely on classical inferential statistics in which either a t-test or an ANOVA is conducted. Recently, several approaches have been undertaken to apply algorithms from machine learning to recorded gait data. Considering publications on both techniques, the question comes up if these methods are related to each other quantitatively. The previous sections 4.2 and 4.3 derived and illustrated these relations. In the following, its application is exemplified on a published study in gait analysis. First, the previously described comparison is extended to dependent samples, and then classification rates $\bar{R}(\mathcal{X})$ are estimated from reported values in the article.

4.4.1 Application to Dependent Samples

In social, psychological and medical science, the dependent samples design is preferred over the independent samples design if applicable. In this case, either subjects of each condition are matched on factors or else each subject serves in all of the c ($c \geq 2$) experimental conditions. The latter one is also referred to the within-subjects or repeated-measurements design. Statistical analysis considers paired, i. e. dependent, samples. In this way, variations of the dependent variable caused by individual properties of each subject have less influence on the test statistics. For validity of the statistical test, the experimenter should take control over order effects into account, e.g. by counterbalancing, for the within-subjects design and random assignment of the subjects to the conditions.

The following description is based upon the textbooks [136, 143]. Generally, the t-test for two dependent samples is applied to interval or ratio data. It is based on the following assumptions:

- Random and independent selection of each sample from the population it represents,

- Homogeneity of variances: $\sigma_1^2 = \sigma_2^2 = \sigma^2$ where σ_1^2 refers to the variance of the population under condition 1 and σ_2^2 under condition 2,
- Samples originate from Gaussian distributions, i.e. $\Xi_i \sim N(\mu_i, \sigma_i^2)$,
- Equal sample size $n_1 = n_2 = n$.

Defining the difference $D_i = x_{i,1} - x_{i,2}$ for each paired sample i, the computed test statistic becomes

$$\hat{t}_{depend} = \frac{\overline{D}}{s_{\overline{D}}} = \frac{\sum_i D_i}{s_{\overline{D}}} \tag{4.37}$$

with the standard error of the mean difference $s_{\overline{D}}$

$$s_{\overline{D}} = \frac{s_D}{\sqrt{n}} \ . \tag{4.38}$$

The estimated standard deviation of the differences s_D is

$$s_D = \sqrt{\frac{\sum_i D_i^2 - \frac{(\sum_i D_i)^2}{n}}{n-1}} \ . \tag{4.39}$$

Thus, analyzing the paired differences results in an one-sample t-test. The hypothesis H_0 is rejected if

$$\hat{t}_{depend} > t_{df,1-\alpha/2} \quad \text{with} \quad df = n - 1 \ . \tag{4.40}$$

The degree of freedom df is reduced by half comparing to a t-test for independent samples. It is noted in [136, p.447], that pairwise data analysis compensates for low df. Pairwise samples are to be preferred if

$$\left[\frac{n(2n+1)}{(n+2)(2n-1)}\right]\left[\frac{(n-1)s_s^2 + ns_D^2}{(2n-1)s_D^2}\right] > 1 \tag{4.41}$$

with

$$s_s^2 = \sqrt{\frac{\sum_i (x_{i,1} + x_{i,2})^2 - (\sum_i (x_{i,1} + x_{i,2}))^2/n}{n-1}} \ . \tag{4.42}$$

Generally, the t-test for paired samples has higher power to detect an effect and is more informative because extraneous effects are more easily to screen out. On the counterpart, it is more sensitive to violence against the homogeneity of variance assumptions than the t-test for two independent samples.

The effect size can be either computed with the pooled standard deviation s_e or with the standard deviation of the differences s_D. The latter one is calculated by

$$d_{depend} = \frac{\bar{D}}{s_D} = \hat{t}_{depend}\sqrt{\frac{2}{n}} \ . \tag{4.43}$$

Dunlap *et al.* in [44] argue that the effect size d_{depend} overestimates the actual effect size

and recommend a corrected formula for estimation of d

$$d = \hat{t}_{depend}\sqrt{\frac{2(1-r)}{n}} \ , \tag{4.44}$$

where r is the correlation coefficient between the two groups. Eq. 4.44 is equivalent to Eq. 4.6. Thus, the relations between d and the classification rate are the same for a dependent t-test if d is reported. Otherwise, d_{depend} needs to be converted to $d = d_{depend}\sqrt{(1-r)}$. The latter case requires reported r values.

The analogue for comparing more than two means for a repeated measures design is the ANOVA for dependent samples. In this case, the total variability SS_T/df_T in the data is divided into between-conditions variability SS_{BG}/df_{BG}, the between-subjects variability SS_{BS}/df_{BS} and the residual variability SS_{res}/df_{res} with

$$\begin{aligned} SS_T &= SS_{res} + SS_{BG} + SS_{BS} \\ &= \sum_{i,j}(x_{i,j} - \bar{x}_i - \bar{x}_j + \bar{x})^2 + n \cdot \sum_{i=1}^{c}(\bar{x}_i - \bar{x})^2 + c \cdot \sum_{j=1}^{n}(\bar{x}_j - \bar{x})^2 \ . \end{aligned} \tag{4.45}$$

The $\hat{F}$ statistic for dependent samples is calculated as follows:

$$\hat{F} = \frac{MS_{BG}}{MS_{res}} = \frac{\frac{1}{c-1} \cdot n \cdot \sum_{i=1}^{c}(\bar{x}_i - \bar{x})^2}{\frac{1}{(n-c)(c-1)} \cdot \sum_{i,j}(x_{i,j} - \bar{x}_i - \bar{x}_j + \bar{x})^2} \ . \tag{4.46}$$

The effect sizes $\eta^2 = \frac{SS_{BG}}{SS_T}$ and partial $\eta_p^2 = \frac{SS_{BG}}{SS_{BG}+SS_{res}}$ are most frequently reported where η_p^2 is usually larger than η^2. Fig. 4.7 (b) can also be used to approximate a classification rate from reported η^2 of a within-subjects analysis because calculation of η^2 equals Eq. 4.12 and $\hat{\eta}^2 \approx \hat{\omega}^2$. For exact approximation, knowledge on MS_{WG} and on the means $\bar{x}_i$ is required to apply e.g. Eq. 4.24.

4.4.2 Selected Application: The Embodiment of Depression and Sadness in Gait

Michalak *et al.* analyze in [106] how gait patterns are associated with sadness and depression. For a number of $n_1 = 14$ participants suffering from a diagnosis of major depression and a number of $n_2 = 14$ never-depressed control participants, differences of five gait features are investigated. Means and standard deviations for each dependent variable are reported. Furthermore, a dependent t-test was applied to the recorded data of each gait parameter. The authors illustrate in a bar plot that all effect sizes are larger than 0.8. For better approximation of corresponding classification rates, the exact value of d is calculated from reported $\bar{x}_i$ and s_i. Tab. 4.4 summarizes the reported measures and the results for the approximated classification rates. A rough approximation of the corresponding classification rates by reading out d from the bar plot and using Fig. 4.5 (a) for approximation would also be possible. Classification rates range between 65% and 74%. The authors note that the results of the inferential analysis do not allow conclusions about the specificity of the gait patterns. This means that, for a classification task, no conclusions can be drawn

Gait Parameter	$\bar{x}_1$	s_1	$\bar{x}_2$	s_2	$\hat{t}_{depend}$	$\hat{t}$	d	$\bar{R}(\mathcal{X})$
Speed	1.07	0.22	1.30	0.17	3.16	3.10	1.17	72%
Arm Swing	274.07	97.62	370.51	82.26	2.69	2.82	1.07	70%
Lateral Body Sway	49.03	16.11	35.63	7.91	-3.59	2.79	1.06	70%
Slumped Posture	1.97	2.99	-3.17	4.93	-3.57	3.34	1.26	74%
Vertical Movement	33.13	13.00	41.10	6.65	2.26	2.04	0.77	65%

Tab. 4.4: Approximation of $\bar{R}(\mathcal{X})$ for the discrimination between walking styles of depressed and non-depressed participants.

Gait Parameter	$\bar{x}_{sad}$	s_{sad}	$\bar{x}_{happy}$	s_{happy}	$\hat{t}_{depend}$	$\hat{t}$	d	$\bar{R}(\mathcal{X})$
Speed	0.84	0.13	1.06	0.17	8.79	4.93	1.45	77%
Arm Swing	218.77	92.43	352.86	139.36	7.97	3.85	1.13	71%
Lateral Body Sway	31.87	12.04	28.51	10.55	-2.69	1.01	0.30	56%
Slumped Posture	-3.25	6.08	-5.89	5.24	-3.37	1.63	0.48	60%
Vertical Movement	27.51	5.16	38.45	8.93	8.62	5.08	1.50	77%

Tab. 4.5: Approximation of $\bar{R}(\mathcal{X})$ for the discrimination between sadness and happiness.

on the confusion of depressed gait styles with other gait styles, e.g. caused by weight or complaints. This is a major issue, because high specificity is required for generalization of approaches in machine learning.

The authors further discuss that depressed and never-depressed participants differ in weight and taking of antidepressant medication that might influence the participant's walking styles. Thus, they carried out a further study where the gait of 23 participants was recorded after mood induction. Considered states are happiness and sadness. Results are listed in Tab. 4.5. Again, the exact effect size d is calculated and the classification rate $\bar{R}(\mathcal{X})$ is approximated. Based on the reports in [106], a classification task based only on observing speed or vertical movement of the body would already discriminate the two emotional states happiness and sadness with 77% accuracy.

4.5 Summary

Computer scientists work together with researchers in biomechanics and medicine within the interdisciplinary research field of gait analysis. This interdisciplinary orientation leads to the application of different methods from statistics to analyze gait data. Researchers in biomechanics and medicine apply inferential statistics, traditionally NHST, to investigate whether the independent variable under investigation causes variations in gait. On the other hand, researchers in computer science and engineering study the prediction of characteristics from observing the gait. This chapter aims to compare these two approaches to analyze gait data. It elaborates dissimilarities, similarities, and quantitative relations between selected methods from inferential and predictive statistics.

Inferential statistics is subdivided in frequentist inference and Bayesian inference. Even

though the latter one is a more flexible approach, frequentist inference is commonly applied in gait analysis. The t-test is the technique for comparing two means and the ANOVA is the technique for more than two means. Both methods assume normally distributed data and homogeneous variances. In machine learning, a Bayes classifier provides the minimum error rate $P_E^\star$ if the distributions of the populations are known. A Bayes classifier with a linear decision border is chosen for the intended comparison because it models the data by Gaussian distributions and assumes a common variance. In the following, the results are summarized for the comparison of the calculated statistics for a t-test or ANOVA and the classification rate of a linear Bayes classifier.

First, the investigated methods from inferential statistics draw conclusions about population parameters whereas predictive statistics aims to assign class membership to single samples. Still, calculated statistics and effect sizes can be converted to the classification rate $\bar{R}$ of a Bayes classifier. Cohen describes in [27] that the effect size d of the t-test can be visualized by the classification rate $\Phi(\frac{d}{2})$. From this follows that $\bar{R}(\mathcal{X}) = \Phi(\frac{\hat{t}}{2n})$ for the $\hat{t}$ statistic. This work additionally derives the relation between statistics of the ANOVA and the classification rate $\bar{R}$, see Eqns. 4.23 - 4.26. A special case of these equations is for two means. In this case, the equations simplify to the relations obtained for the t-test. In the context of interdisciplinary research, these equations facilitate that classification rates $\bar{R}$ of a linear Bayes classifier can be estimated from reported statistics of frequentist inference. In doing so, a range of possible classification rates can be given for a research problem without the need to have access to the recorded data which may not be public available.

Second, these relations are illustrated for population parameters. Possible classification rates $\bar{R}$ can be approximated graphically from these plots. Furthermore, they show that $\bar{R}$ is lower than 66% for reported small and medium effect sizes of the t-test and the ANOVA. The relation between statistics of the ANOVA and the classification rate $\bar{R}$ depends further on the distribution of the means. Upper and lower bounds are derived for this dependence in the Eqns. 4.27 - 4.36. The bounds differ slightly if they are calculated for the parameters μ_i and σ^2 of the population or for their estimates $\bar{x}_i$ and s^2.

Third, statistics such as $\hat{t}$ and $\hat{F}$ are used for feature selection in machine learning. If the assumptions of the t-test are valid, sorting features after $\hat{t}$ equals sorting features after $\bar{R}$. This is not the case for sorting features after $\hat{F}$, because the relation between $\bar{R}$ and statistics of the ANOVA additionally depends on the distribution of the means, as illustrated in Fig. 4.9. Guo and Nixon mention in [58] that one limitation of using $\hat{F}$ for feature selection is a missing explicit and definite relation with classification accuracy. This limitation is overcome by the Eqns. 4.23 - 4.26 and the upper and lower bounds given by the Eqns. 4.27 - 4.36.

Finally, the application of the relation between inferential and predictive statistics is exemplified for an article about the embodiment of sadness and depression in gait. After an extension to dependent samples, classification rates from reported statistics are estimated.

In conclusion, classification rates can be estimated from reported statistics of frequentist inference which compare the difference between means. This concept is generalizable to applications in which both inferential and predictive data analysis is of interest.

4.6 Limitations

The comparison within this chapter focuses on univariate techniques. In machine learning, classification rates improve if not only a single feature is used. Thus, a future direction is towards comparing multivariate methods. Furthermore, the relation between statistics of the investigated techniques from frequentist inference and the classification rate of a linear Bayes classifier assume that the data are normally distributed, that the distributions have homogeneous variances, that the sample size is equal for all groups, and that the independent variable is divided in few, discrete levels. Therefore, analyzing relations for nonparametric or regression techniques are a further future direction.

It should be additionally noted that recognition rates calculated by applying leave-one-out or cross-fold validation is preferred over classification of all samples in machine learning. Yet, this requires access to the complete data set and, hence, these recognition rates cannot be estimated from reported statistics alone.

Bayesian inference is a more flexible approach than frequentist inference in inferential statistics. For the special case of comparing two means and assuming a common variance, the same $\hat{t}$ statistic is computed and the relation between $\hat{t}$ and $\bar{R}$ can be applied. This chapter concentrates on techniques from frequentist inference and considers only the techniques from Bayesian inference which are analogous to the t-test and the ANOVA.

Finally, the application of both inferential and predictive statistics depends on the application. The area of intersection has been defined in section 4.1.1. If the relation between the dependent variable and the observations is obvious in an application, applying inferential statistics is not required. On the other hand, predictive analysis is needless if the task is only to verify whether there is a difference in the means. Application of both methods is only useful if predictive analysis is in the focus but the relation between cause and observation is not obvious.

5 Recognition of Affect in Gait Patterns

Nonverbal communication plays a major role in future robotics to enhance natural human-robot interaction. Within this research goal, affective computing faces the challenge to automatically recognize a human's affective state. Detection of affect is based on observing facial expressions, linguistic as well as acoustic features in speech, physiological parameters, gesture, and body motions [122, 123, 161]. Considering especially body motions, psychological studies indicate that affective states are also expressed in the way people walk, see the review in chapter 3.1 and also the studies about human perception in chapter 3. As each modality has its limitations, recognition based on combining different modalities seems to be more reliable for real-world applications [57, 161]. To provide an additional modality and to enhance recognition of affect at distance, the human gait is studied in terms of its ability to reveal a person's affective state. Further possible applications for predicting affective states from gait are the cognitive household and high-security locations, e.g. airports.

Only one study has yet investigated the recognition of emotions from gait patterns with techniques from machine learning [69]. In this case, recognition rates are above chance. Furthermore, the previous chapter 4 provides a method to estimate classification rates from reported test statistics and, for distinguishing sadness and happiness in gait, rates of up to 79% are approximated from reported test statistics [106] in section 4.4.

Within this chapter, machine learning algorithms are developed therefor. The investigations especially focus on 1) the influence of an individual's walking style on emotion recognition, 2) the emotion model on which the classification task relies, and 3) on the recognition of emotions in gait versus other attributes such as identity or gender. The motion gait is characterized by many degrees of freedom so that it is often assumed that the gait is as individual as one's fingerprint. Recording gait patterns provides high-dimensional data sets; therefore efficient dimension reduction techniques are required for successful classification. Within this chapter, the unsupervised techniques principal component analysis (PCA) and kernel PCA (KPCA) are compared with the supervised techniques linear discriminant analysis (LDA) and general discriminant analysis (GDA) for dimension reduction. Furthermore, catching the temporal information of gait trajectories with PCA is compared with extracting statistical parameters. Although expression of affect during walking is covered by the primary task of locomotion, recognition of affective states has been accomplished based on observation of a single stride. A further result is that especially affective states which differ in arousal are suitable to be detected in gait patterns.

The walking style is also influenced by other factors. Therefore, a comparative study on different, marker-based gait databases is conducted. Either identity, gender, exhaustion, or affective states are recognized depending on the available recordings in the gait databases. The identity of the walker is best recognizable within these factors. As the dimension of features is high and only a small number of training samples are available, the within-

scatter matrix $\boldsymbol{S}_W$ becomes singular when applying LDA. A mathematical proof is derived that in this case a support vector machine (SVM) with a Gaussian kernel calculates the same recognition rate as a hard-margin classifier or nearest neighbor classifier.

The remainder of this chapter is organized as follows: Section 5.1 summarizes the state of the art on marker-based gait analysis and emotion recognition from motions. Methods for dimension reduction and classification are described in section 5.2. Section 5.3 presents the results on the comparison of different pattern recognition algorithms applied to various aspects in emotion recognition. This analysis is conducted for the Munich database. This section 5.3 has been previously published to a large extend in [170]. Afterwards, different, marker-based gait databases are analyzed in section 5.4. The recognition of affective states based on analyzing gait recordings is compared with the recognition of the identity of the walker, gender, and exhaustion. Finally, a conclusion is given in section 5.5 and limitations are discussed in section 5.6.

5.1 Pattern Recognition

Pattern recognition, also referred to as machine learning in computer science, deals with the extraction of knowledge from sensor data. This task is divided in data preprocessing, feature extraction, and classification. A learning algorithm can be either supervised, where the training data is labeled, or unsupervised. The textbooks of Abe [2], Bishop [9], Duda [43], and Tan *et al.* [147] provide an excellent description about pattern recognition techniques. The summary of the methods used within this chapter is based on these textbooks. In general, pattern recognition tasks have the following challenges in common [43]:

- The *No Free Lunch Theorem* says 'that for both static and time-dependent optimization problems, the average performance of any pair of algorithms across all possible problems is identical' [155]. Thus, superior performance of an algorithm is related to the problem under investigation and its characteristics, e.g. prior information, data distribution, and amount of training data.

- Similarily, the *Ugly Ducking Theorem* says that there is no problem-independent best set of features.

- In a high-dimensional feature space, classification functions have the potential to be much more complicated to determine than in a low-dimensional feature space, e.g. the estimation of a one-dimensional density function requires less samples than a high-dimensional density function. This is commonly known as the *Curse of Dimensionality*. Thus, the aim is to incorporate knowledge about the data and to reduce the dimensionality.

- The trade-off between *Bias and Variance* considers the quality of a model which is estimated on various data sets of the same population. If a model has many parameters, it will fit the data well but variance for the quality of fit will be high. A model with few parameters will result in a lower variance but high bias. A low generalization error is desirable in machine learning; therefore, low variance is more

important than low bias for a classification task. A large training size and prior knowledge on the model lowers both criteria.

Facing these challenges, several techniques from pattern recognition are compared in the following for recognition of affect in gait patterns. As marker-based gait recordings are high-dimensional, focus lies in particular on dimension reduction. This simultaneously avoids the curse of dimensionality and reduces the number of parameters trained for the model.

5.1.1 Marker-based Gait Analysis

Although recognition using vision-based features extracted from the shape of a walker is more applicable for real-world scenarios, model-based features give more insights in the underlying kinematics. For this reason, the latter one is preferred in clinical, biomechanical, and methodological studies. Advantages of high resolution and more reliable calculation of kinematic parameters faces an artificial setup required for marker-based optical motion tracking.

Recordings of gait patterns are characterized by high dimensionality, temporal dependency, high variability, and nonlinearities [21, 22]. Furthermore, gait analysis lacks the definition of a universal normal gait pattern, because an individual's gait is influenced by many factors like age, weight and complaints. This makes the gait suitable for biometric identification in computer vision [10, 71, 117, 139], but complicates recognition of other factors.

Marker-based gait analysis is commonly used in medical and biomechanical studies for more than a decade. Still, data analysis is based traditionally on statistics. To improve the extraction of useful information from highly correlated time-dependent gait parameters, several approaches have been undertaken, including methods from machine learning [21, 22, 69, 131, 150, 156]. An overview is given in Chau's reviews [21, 22]. Clustering, multivariate techniques, artificial neural networks (ANN), and time-frequency analysis have been applied to short-term recordings. Fractal analysis is only applicable to long-term recordings, as it estimates long-range correlations. However, Chau points a lack of objective comparisons between multiple methods out. Wu and Liu investigate the capability of KPCA to capture nonlinear relationships in gait patterns [156]. KPCA slightly increases recognition of age in comparison to PCA. Troje presents a two-stage PCA for gender recognition [150]. Roether *et al.* introduce a phase-adapted blind source separation algorithm to minimize redundancy in the parametrization of gait patterns [131]. Still, the comparison regarding the number of kinematic parameters considered for analysis, different methods to capture the temporal characteristics of gait, and different classifiers is an open issue.

Applications that are most frequently studied in gait analysis are therapeutic support for patients with gait complaints, gait-based human identification, and discrimination between human motion types [10, 21, 22, 71, 86, 91, 117, 139]. Besides further improvement in human motion reconstruction from recorded data, fundamental advances in behavior representation of motions is required for a wider range of applications for human movement analysis [109].

5.1.2 Recognition of Emotions in Movement

Evidence from psychology supports that emotions are not only expressed in facial expressions and speech, but also in posture and body movements. Still a debate is ongoing on the specificity of the expressiveness. Ekman and Friesen stated in [47] that bodily cues express more the quantity than the specificity of emotions, whereas other studies support that also emotion-specific movements and posture characteristics seem to exist [31, 153]. From the engineering point of view, several studies have investigated the detection of emotions from body movements. An overview is given in Tab. 5.1. Studies differ with regard to emotional states, methodology, and whether gestures, postures, or general motions are investigated.

Bernhardt and Robinson especially focus on non-stylized body motions [7]. Stylized and non-stylized motions differentiate in their primary task. The primary task of stylized motions is expressiveness, where non-stylized motions can be actions such as knocking or walking. Furthermore, they distinguish between person-dependent and inter-individual recognition. This is motivated by a principle of character animation. It says that one character would do an action differently in two different emotional states, and that two characters would not do an action in the same way [89]. Taking into account that individuals may knock differently, they report an improvement of 31% in recognition rate. In particular, the affective states angry and sad are better recognizable than neutral or happy.

Kleinsmith *et al.* examine the role of affective dimensions in static postures for automatic recognition [79]. Error rate is lower than 21% for each affective dimension arousal, valence, potency, and avoidance based on an ANN. They conclude that the dimensional approach can be seen as a comparable alternative to discrete affective categories for analyzing affect in postures.

Kapur *et al.* compare in [73] the performance of different classifiers, namely decision tree, logistic regression, naive Bayes, support vector machine (SVM), and ANN for stylized motions. Recognition rate ranges between 86% and 93% for person-dependent and between 62% and 84% for inter-individual recognition. In addition, Meservy *et al.* analyze nonverbal behavior for deception detection in the context of national security [105]. In their study, they compare the performance of discriminant analysis (55%), decision tree (58%), ANN (71%), and SVM (71%). Even though recognition rates are above chance, the authors note that still many technical and utilization challenges need to be overcome. Among others are detection across individuals and varying level of truthness.

Gunes and Piccardi combine recognition based on upper body gestures with facial expressions [57]. Their results, gained on the FABO database, show that recognition from fused face and body modalities performs better than from the face or body modality alone. For a random forest classifier, recognition rates are 34% for facial expressions, 77% for gestures, and 84% for feature-level fusion of both modalities. They reason that it is easier for a vision-based system to model and recognize affect from global body and head motions than from atomic movements in the face.

Author	Year	Movement	Affective States	Feature Extraction	Classifier	Recog. Rate
Bernhardt *et al.* [7]	2007	Non-stylised knocking [98]	Neutral, happy, angry, sad	Energy function	SVM	50%, 81%⋆
Gunes *et al.* [57]	2009	Gesture (upper body)	12 affective states	PCA	Random forest	77%
Janssen *et al.* [69]	2008	Gait	Normal, happy, sad, angry	Kinetic data	ANN	84%⋆
Kapur *et al.* [73]	2005	Stylized body movements	Sad, joy, anger, fear	Mean of velocity and acceleration, standard deviation of position, velocity and acceleration	ANN	85%, 93%⋆
Kleinsmith *et al.*[79]	2007	Static posture	Valence, arousal, potency, avoidance, 3 discrete states for each dimension	24 selected features	ANN	79−81%
Meservy *et al.* [105]	2005	Upper body motions	Guilty, innocent	Independent samples t-test	SVM, ANN	71%

Tab. 5.1: Recognition of affective states from posture and movements (⋆ indicates person-dependent rates).

Only one work studies recognition of affect in walking with techniques from machine learning [69]. Janssen *et al.* investigated the recognition of four affective states by means of ANNs. Person-dependent recognition rate reaches 84% in average based on kinetic data measured with a force plate in the ground. However, inter-individual recognition rate remains around chance level. Walking during listening to calming, excitatory or no music is to 79% classifiable using kinematic parameters.

Even though first studies on the recognition of emotions from bodily movements have been conducted, recognition of emotions from bodily movements is still a relatively unexplored field comparing to a numerous number of publications on facial expressions and speech. Reviews on affect detection suggest that further research is required both in terms of categorical and dimensional emotion definition [18, 56]. This is a challenging task because a psychological framework – similar to the facial action coding system – does not yet exist for expressive body movements. Furthermore several approaches from machine learning have been undertaken, where best results are gained with SVMs or ANN. Features are predominantly selected by the researchers to limit the feature space with respect to a limited amount of training data. The two approaches which have been applied from machine learning are PCA and an independent samples t-test.

5.2 Methods

Classification of gait patterns is challenging for techniques from machine learning, because data of gait recordings are high-dimensional, time-dependent, highly variable, and gait variables interact in a nonlinear relationship [21]. Furthermore, retrieving the affective state of the walker from gait patterns faces the challenge to extract patterns which are not directly observable but rather covered by the primary task of locomotion. Fig. 5.1 illustrates the influence of affect on the right shoulder angle. The trajectories, which describe for- and backward movement of the right arm, are from the same walker as in Fig. 3.4. Influence of affect on walking differs between individuals as well as between joint angles and is usually less distinguishable by graphical investigation as in Fig. 5.1. Since walking is a symmetric movement, following analysis focuses on the right side of the body. Furthermore, joint angles are used as a basis for feature extraction and classification, so that comparing performance of automatic classification with human observers is based on the same information obtained from the recordings.

Improvement in accuracy can be achieved by either optimizing the classifier or feature extraction. This section focuses on comparing three standard classifiers and improving feature extraction so that accuracy of automatic classification matches or even outperforms human evaluation. Two approaches are compared to extract features from the joint trajectories. The former one is based on statistical parameters in combination with different dimension reduction techniques. Dimension reduction is subdivided in projection (e.g. PCA), compression (e.g. using information theory), and selection methods. For this application, feature space projection and feature selection are applied. The latter approach applies PCA directly to the trajectories over time to extract temporal information.

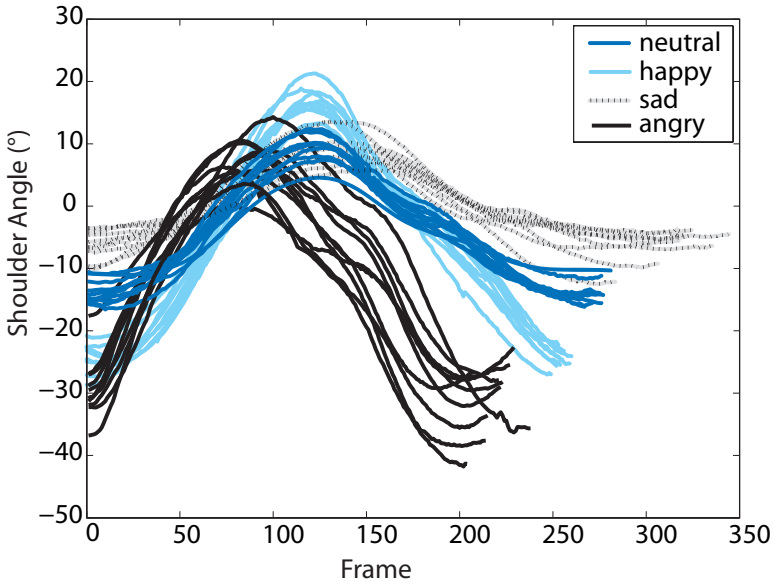


Fig. 5.1: Trajectories of the right shoulder angle differ with respect to expressed affect. 10 recordings of one walker are plotted for each affective state.

5.2.1 Statistical Parameters of Joint Angle Trajectories

In literature, the expression of affective states in movement is often associated with its velocity [32, 126, 131]; therefore the parameters stride length, cadence, and velocity are analyzed separately. Minimum, mean and maximum are calculated for each joint angle. Considered joint angles are head, neck, shoulder, elbow, thorax, spine, pelvis, and foot progress angle for each rotation axis. Furthermore, humans observe more the upper part than the lower part of the body during an emotion recognition task, see chapter 3.3.2. This suggests to select a subset of joint angles from the upper body part. Thereby, the joint angles, see Tab. 5.2, are noted in literature to be influenced significantly by affect [32] and form a subgroup which is analyzed separately. Summing up, the kinematic parameters are split in three groups. Classification and optional dimension reduction are based on $\mathbf{x}_{kin} \in \mathbb{R}^d$ with dimension M, containing one of the following sets of kinematic parameters

- Solely velocity, stride length and cadence (VSC), $d = 3$,
- Minimum, mean and maximum of significant joint angles including VSC, $d = 15$,
- Minimum, mean and maximum of all joint angles including VSC, $d = 69$.

Transforming and reducing the feature space can improve classification. With this purpose, PCA, KPCA, LDA and GDA are applied to the significant subsection and to all joint angles [9, 43, 151]. Mean and standard deviation of the feature vector $\mathbf{x}_{kin,norm}$ are normalized for PCA and KPCA.

Factor	Parameter
Stride Length	Length of One Stride
Cadence	Time for One Stride
Velocity	Cadence/(Stride Length)
Neck Angle (Forward Tilt)	Min, Mean, Max
Shoulder Angle (Flexion)	Min, Mean, Max
Shoulder Angle (Abduction)	Min, Mean, Max
Thorax Angle (Forward Tilt)	Min, Mean, Max

Tab. 5.2: Kinematic features of significant subsection.

Principal Component Analysis (PCA)

PCA is also known as Karhunen-Lòeve transformation. It calculates the eigenvectors, further on called principal components (PCs), of the covariance matrix and then transforms the original data space to an orthogonal set of principal components. Principal components $\mathbf{u}_i$ with highest eigenvalues λ_i represent the vectors with maximum variance in the data set. The eigenvalue problem to solve is defined as

$$\left(\frac{1}{N} \sum_{j=1}^{N} \mathbf{x}_{kin,norm,j} \mathbf{x}_{kin,norm,j}^{T} \right) \mathbf{u}_i = \lambda_i \mathbf{u}_i \quad \text{with } i = 1, ..., d \ , \tag{5.1}$$

with N observations of $\mathbf{x}_{kin,norm}$. Original data is mapped on up to a maximum of d principal components. Dimension reduction is achieved if the coefficients of the first m principal components are used for classification with $m < d$. Since PCA is an unsupervised technique, it is not guaranteed that the projection which maximizes the variance in the data also maximizes the representation of affect in the transformed feature space. Furthermore, PCA is a linear technique and does not take underlying nonlinearities into account.

Kernel Principal Component Analysis (KPCA)

A non-linear extension of PCA is KPCA [142]. Its principle is that it first maps the data to a more suitable feature space $\mathbb{R}^m$ and then applies PCA in $\mathbb{R}^m$. This involves two advantages. On one hand, an appropriate mapping can enhance the separability of the clusters. And on the other hand, the dimension m can be larger than d so that the number of possible eigenvectors increases to N. Given such a map $\boldsymbol{\phi}$ which can be nonlinear

$$\boldsymbol{\phi} : \mathbb{R}^d \rightarrow \mathbb{R}^m, \quad \mathbf{x}_{kin,norm} \mapsto \mathbf{X}_{kin,norm} \ ,$$

each sample $\mathbf{x}_{kin,norm}$ is mapped to $\mathbf{X}_{kin,norm}$ in $\mathbb{R}^m$. Up to this point explicit knowledge about $\boldsymbol{\phi}$ is required. Applying PCA in $\mathbb{R}^m$, means to solve $\lambda_i \mathbf{u}_i = \mathbf{C}\mathbf{u}_i$ with

$$\mathbf{C} = \frac{1}{N}\sum_{j=1}^{N} \boldsymbol{\phi}(\mathbf{x}_{kin,norm,j})\boldsymbol{\phi}^T(\mathbf{x}_{kin,norm,j}) \ . \tag{5.2}$$

Schölkopf *et al.* show in [142] that this eigenvalue problem can be reformulated as

$$N\lambda\mathbf{a} = \mathbf{K}\mathbf{a} \quad \text{with} \quad k_{ij} := \boldsymbol{\phi}^T(\mathbf{x}_{kin,norm,i})\boldsymbol{\phi}(\mathbf{x}_{kin,norm,j}) \ , \tag{5.3}$$

where the dot product $\boldsymbol{\phi}^T\boldsymbol{\phi}$ is substituted by a symmetric kernel function $k(\cdot,\cdot)$. It can be shown that if the conditions of a Mercer's kernel are true for $k(\cdot,\cdot)$, always a mapping $\boldsymbol{\phi}$ exists. Applying the kernel trick does not only reduce computational complexity but also has the advantage that explicit knowledge about the mapping $\boldsymbol{\phi}$ is not required. Within this study, a polynomial kernel $k(\mathbf{x},\mathbf{x}') = (\mathbf{x}^T\mathbf{x}')^D$ and a Gaussian kernel $k(\mathbf{x},\mathbf{x}') = \exp((-\|\mathbf{x}-\mathbf{x}'\|^2)(2\sigma^2)^{-1})$ are used.

Linear Discriminant Analysis (LDA)

In contrast to algorithms based on PCA, LDA considers the class membership for projection. The concept of LDA is to maximize the separability of the class means and to minimize the variance around these means. The eigenvectors $\mathbf{w}_k$ which span the lower-dimensional space after projection are a linear combination of the original dimensions. For a c-class problem, the maximum number of eigenvectors is $(c-1)$. Based on the total mean $\mathbf{m} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_{kin,i}$, the mean $\mathbf{m}_j$ for samples $\mathbf{x}_{kin,i,j}$ of each class j, and the number of samples n_j for each class, the between-class scatter matrix $\boldsymbol{S}_B$

$$\boldsymbol{S}_B := \sum_{j=1}^{c} n_j(\mathbf{m}_j - \mathbf{m})(\mathbf{m}_j - \mathbf{m})^T \ , \tag{5.4}$$

and the within-class scatter matrix $\boldsymbol{S}_W$

$$\boldsymbol{S}_W := \sum_{j=1}^{c}\sum_{i=1}^{n_j}(\mathbf{x}_{kin,i,j} - \mathbf{m}_j)(\mathbf{x}_{kin,i,j} - \mathbf{m}_j)^T \ , \tag{5.5}$$

are defined. A measure for maximizing the between-class separability while minimizing the within-class separability is $\frac{\det|\boldsymbol{S}_B|}{\det|\mathbf{S}_W|}$. The advantage of this measure is that it is maximized for a projection matrix which contains the eigenvectors of $\mathbf{S}_W^{-1}\boldsymbol{S}_B$. Thus, the generalized eigenvalue problem is

$$\boldsymbol{S}_B\mathbf{w}_k = \lambda_k\boldsymbol{S}_W\mathbf{w}_k \quad \text{with } k = 1,...,c-1 \ . \tag{5.6}$$

At least $d+c$ samples are required so that $\boldsymbol{S}_W$ is not singular and that the eigenvectors are the optimum solution for maximizing $\frac{\det|\mathbf{S}_B|}{\det|\boldsymbol{S}_W|}$. If $\boldsymbol{S}_W$ is singular, either a combination of PCA and LDA or advanced algorithms for solving the eigenvalue problem can be used.

Regarding the latter, both [39] and [96] show that applying the generalized Schur (QZ) decomposition for solving the eigenvalue problem leads to similar or better results than related approaches even in the case of the small sample size problem. The QZ algorithm is a standard implementation in MATLAB and used within this study. Furthermore, Martinez and Kak note that PCA can perform better than LDA even though it does not take class affiliations into account [99].

Similarly to KPCA, General Discriminant Analysis (GDA) performs LDA in a high-dimensional feature space [5]. Again, using a kernel avoids explicit mapping into the high-dimensional space. Polynomial as well as Gaussian kernels have been applied to calculate inter-individual and person-dependent recognition rates. Accuracy does not exceed chance level, so that results are not further reported.

5.2.2 Modeling Joint Angle Trajectories by Eigenpostures

In contrast to computing the minimum, mean and maximum of a time series that corresponds to a joint angle, the approach based on eigenpostures directly applies PCA to the complete data set of one gait. The procedure, see Fig. 5.2, is lent on eigenpostures and eigenwalkers, as proposed by Troje [150].

First, spatial information in the data set of one gait is reduced. The vector $\mathbf{x}(t) \in \mathbb{R}^{23}$ contains x, y and z-rotation of each joint angle at time step t. The complete data set of one walk is described by the matrix

$$\mathbf{X} = [\mathbf{x}(1)\ \mathbf{x}(2)\ ...\ \mathbf{x}(T)]\ ,$$

containing T frames. It is required for PCA that the matrix $\mathbf{X}$ is normalized. Therefore, the mean posture $\mathbf{p}_{mean} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}(t)$ is subtracted, and the data set is divided by its standard deviation $\mathbf{p}_{std}$ for unit variance. PCA is applied to $\mathbf{X}_{norm}$. Dimension reduction is achieved by using only the four eigenvectors, called eigenpostures $\mathbf{p}_j$, with highest eigenvalues for further data analysis. Hence, the gait of one walker $\mathbf{p}(t)$ at time step t is described by

$$\mathbf{p}(t) = \mathbf{p}_{mean} + diag(\mathbf{p}_{std})\sum_{j=1}^{4} c_j(t)\mathbf{p}_j\ ,$$

where $c_j(t),\ 1 \leq j \leq 4$, are the coefficients of $\mathbf{x}_{norm}(t)$ after transformation.

Temporal information is covered by the coefficients $c_j(t)$. As human walking is periodic and each recording covers at least one stride, Fourier transformation (FT) is applied to model the temporal behavior of the coefficients $c_j(t)$. The leakage effect is reduced using a Hamming window. Main frequency f_j and phase Φ_j are extracted. The phase of the first eigenposture Φ_0 is set to zero and Φ_j for $j = 2, 3, 4$ are shifted by their difference to Φ_0. The individual gait of one person $\mathbf{p}(t)$ is modeled as follows:

$$\mathbf{p}(t) = \mathbf{p}_{mean} + diag(\mathbf{p}_{std})\mathbf{p}_1 \sin(2\pi f_1 t) + diag(\mathbf{p}_{std})\sum_{j=2}^{4}\mathbf{p}_j \sin(2\pi f_j t + \Phi_j)\ . \tag{5.7}$$

Third, a second PCA extracts most relevant information among different walks of one

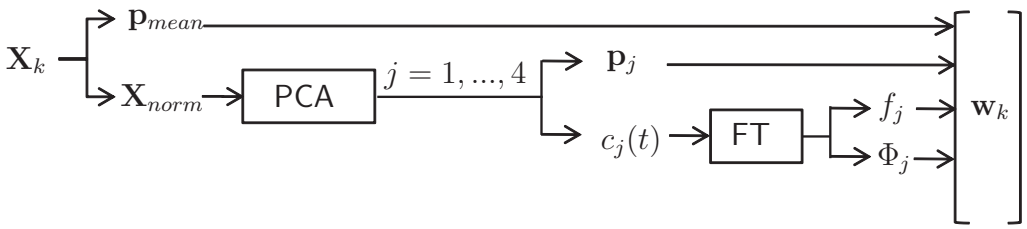


Fig. 5.2: The mathematical description $\mathbf{w}_k$ of one walk k contains the mean posture $\mathbf{p}_{mean}$, 4 eigenpostures $\mathbf{p}_j$, 4 frequencies f_j and 3 phase shifts ϕ_j.

walker and within different walkers. The description of a single walk $\mathbf{w}_k$ contains the mean posture $\mathbf{p}_{mean}$, four eigenpostures $\mathbf{p}_{j,k}$, four frequencies $f_{j,k}$ and three phase shifts $\Phi_{j,k}$

$$\mathbf{w}_k \;=\; [\mathbf{p}_{mean}^T\ \mathbf{p}_{1,k}^T\ ...\ \mathbf{p}_{4,k}^T\ f_{1,k}\ ...\ f_{4,k}\ \Phi_{2,k}\ ...\ \Phi_{4,k}]^T\ . \tag{5.8}$$

The matrix $\mathbf{W}$ contains e.g. all walks $\mathbf{w}_k$ of one walker expressing different affective states for person-dependent recognition. Eigenvectors of the second PCA, which is applied to $\mathbf{W}$, are called eigenwalkers. Their coefficients are used for final classification. Results of this approach, later referred to as PCA-FT-PCA, are given for classification based on the coefficients of all eigenwalkers.

It is noted, that replacing either the first PCA, the second PCA or both PCAs by KPCA does not improve accuracy significantly [166]. Furthermore, the second PCA can be replaced by LDA or GDA [179]. In this case, number of instances in the training set is approximately four times smaller than dimension of a single instance $\mathbf{w}_k$, so that performance of LDA and GDA is low.

5.2.3 Classification

For recognition several standard classifiers are compared. Two parametric classifiers and one non-parametric classifier are applied to the data. Description of the classifiers is primarily based on [2, 9, 43, 87, 147]. The data set contains n_i samples for each class i with a total number of samples $N = \sum_{i=1}^{c} n_i$. As the number of samples in the data sets is small, the recognition rate is calculated using leave-one-out cross validation. Each data point $\mathbf{x} \in \mathbb{R}^d$ contains d features. The true state of nature is denoted as ω_i further on.

Bayes Formula and Bayes Error Rate $\mathbf{P_E^\star}$

Given the class-conditional probability density $p(\mathbf{x}|\omega_i)$ and the prior probability $P(\omega_i)$, the posterior probability $P(\omega_i|\mathbf{x})$ is calculated by Bayes' formula

$$P(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)P(\omega_i)}{p(\mathbf{x})}\ , \tag{5.9}$$

where $p(\boldsymbol{x}) = \sum_{i=1}^{c} p(\mathbf{x}|\omega_i)P(\omega_i)$ is a scaling factor. The probability of error $P_{E,i}$ is

$$P_{E,i} = 1 - P(\omega_i|\mathbf{x})\ .$$

By selecting the ω_i which maximizes $P(\omega_i|\mathbf{x})$, the overall error P_E is minimized. Assuming that the distribution which generates the data is known, the Bayes' error rate $P_E^\star$ is the best performance a classifier can achieve. However, the distribution is normally estimated from the training data for real-world applications and the performance depends on how well the model approximates the real distribution.

Naive Bayes

The central limit theorem says that the sum of arbitrarily, identically distributed random variables can be approximated by a Gaussian distribution. Real-world measurements are often influenced by a large number of random processes. Considering the central limit theorem, it is not so far off that approximation of an unknown distribution with a Gaussian may provide good results in practice.

The general multivariate normal density function is given by

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right] \ , \tag{5.10}$$

where $\boldsymbol{\mu}$ is the mean and $\mathbf{\Sigma}$ the covariance matrix, which is symmetric and positive semidefinite. Thus, $d + d(d+1)/2$ parameters need to be estimated for a d-dimensional feature vector. If the features are statistical independent, i.e., $\sigma_{jk} = 0 \ \forall j \neq k$, $\mathbf{\Sigma}$ reduces to a diagonal matrix. A Naive Bayes classifier assumes statistical independence of the features and $p(\mathbf{x}|\omega)$ simpilfies to $p(\mathbf{x}|\omega) = \prod_{j=1}^{d} p(x_j|\omega)$. If the variances are estimated separately for each class, it is called quadratic otherwise linear. Even though the assumption of conditional independence rarely holds for real data, Naive Bayes often reaches good performance. Explanations are that dependencies among the features may cancel each other out and that for minimum-error classification the absolute value of the posterior probability can be wrongly estimated as long as the maximum probability is assigned to the correct class [163]. Characteristic for a Naive Bayes Classifier is that it is robust to isolated noisy samples and irrelevant features. Furthermore, it can handle missing features. However, correlated attributes can decrease the performance if the dependencies among all features do not work together in a way that they cancel each other out. For further gait analysis, a quadratic Naive Bayes classifier is applied because it can obtain a good estimate of the likelihoods even if the number of training samples is small.

Nearest Neighbor

The principle of the Nearest Neighbor (NN) classifier is that it assigns the class label of the nearest training sample to the test sample. Despite its simplicity, it can be shown that the error probability of a NN classifier P_E is always lower than twice the Bayesian error rate $P_E^\star$ for infinite samples. The error rate P_E is bounded by

$$P_E^\star \leq P_E \leq P_E^\star(2 - \frac{c}{c-1}P_E^\star) \ .$$

NN belongs to the non-parametric techniques because it does not explicitly model the underlying distribution which generates the data. It can produce arbitrarily shaped de-

cision boundaries but is susceptible to noise. Furthermore, any metric can be chosen as distance measure between two samples. The Euclidean distance is used in this work. As a lazy learner, the distances of a test sample to all training samples is calculated when a new training sample comes in; therefore computational effort can become larger than for a parametric technique during the test phase in particular if the feature space is high dimensional and many distances need to be calculated. Furthermore, non-parametric techniques are more susceptible to the curse of dimensionality. The reason therefor is that knowledge about the underlying distribution can not be included in a model and a high density of training samples in the feature space is required for good performance.

Calculation of the distance can be extended to a number of k nearest neighbors and the decision about class assignment is made by a majority vote. This approach can also be viewed as an estimation of $P(\omega_i|\mathbf{x})$. The posterior probability $P(\omega_i|\mathbf{x}) \approx k_i/k$ is estimated based on the k_i nearest neighbors of each class ω_i. Choosing the appropriate number of k nearest neighbors is a trade-off between choosing only the nearest neighbors and as many neighbors as possible for a reliable decision. In the case of infinite samples, the upper bound of the error rate P_E approaches $P_E^\star$ if the number of nearest neighbors k approaches infinity. To study how well class assignment works if only the nearest neighbor is chosen to approximate $P(\omega_i|\mathbf{x})$, k is set equal to 1 for the following analysis of gait patterns.

Support Vector Machine

Besides ANN, Support Vector Machines (SVM) afford good results in emotion recognition, see Tab 5.1. SVMs extend a simple linear decision border in two ways: 1) the optimum hyperplane with respect to generalization for a given data set is found, and 2) the feature space can be mapped to a higher dimensional space. The linear decision hyperplane for a d-dimensional feature space is

$$f(\mathbf{x}) = w_1x_1 + w_2x_2 + ... + w_dx_d + w_0 = \mathbf{w}^T\mathbf{x} + w_0 \tag{5.11}$$

with $\mathbf{w} \in \mathbb{R}^d$. For a two-class problem, the function $y_k = \text{sign}f(\mathbf{x}_k) : \mathbb{R}^d \rightarrow \{-1, 1\}$ gives the class association ω_1 or ω_2 for one sample k. If the samples are linearly separable in the feature space, the number of possible decision hyperplanes is infinite. A maximum margin classifier maximizes the distance $\frac{\mathbf{w}^T\mathbf{x}}{\|\mathbf{w}\|}$ between the training samples and the decision border. In doing so, it is expected to provide best generalization for a given training set. The decision hyperplane is found by the following optimization problem

$$\begin{aligned} &\min_{\mathbf{w}\in\mathbb{R}^d} \|\mathbf{w}\|^2 \\ \text{subject to:} \quad &y_k\ (\mathbf{w}^T\mathbf{x}_k + w_0) \geq 1 \quad \text{for } k = 1...N\ . \end{aligned} \tag{5.12}$$

This is also called a hard margin classifier. This optimization problem has no feasible solution, if the training samples are not linearly separable. To allow margin violations, a

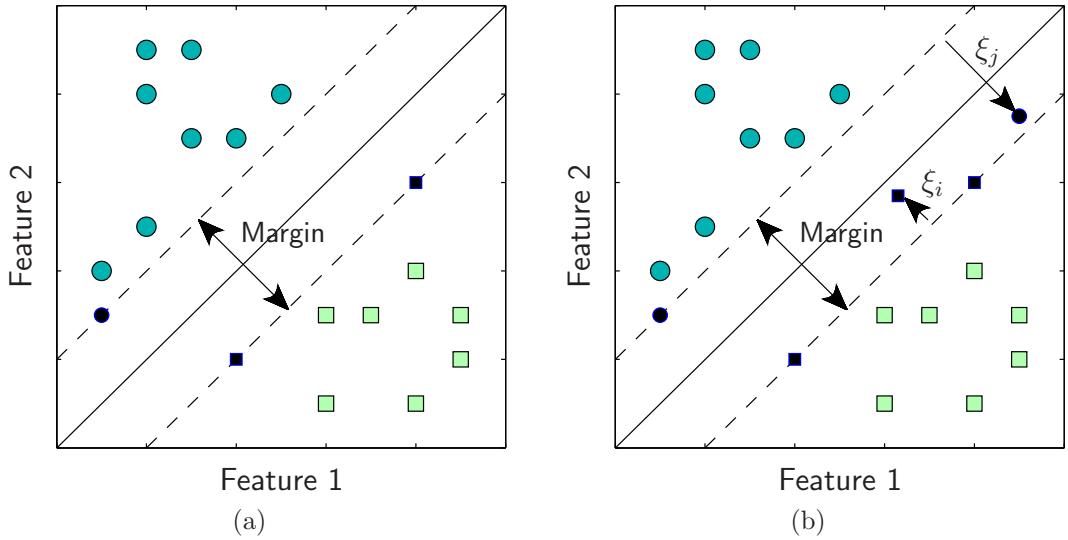


Fig. 5.3: (a) If the data set is linearly separable, the margin is maximized without violations. (b) If the data is not linearly separable, a soft-margin SVM allows violations. Support vectors are illustrated by ■ and ●.

slack variable ξ_k is introduced. The optimization problem becomes

$$\min_{\mathbf{w}\in\mathbb{R}^d,\xi_k\in\mathbb{R}^+} \|\mathbf{w}\|^2 + C\sum_{k=1}^{N}\xi_k^p$$
$$\text{subject to:} \quad y_k\ (\mathbf{w}^T\mathbf{x}_k + w_0) \geq 1 - \xi_k \quad \text{for } k = 1...N\ , \tag{5.13}$$

with the regularization parameter C, and $p = 1$ for a L1 soft-margin SVM or $p = 2$ for a L2 soft-margin SVM. A L2 soft-margin classifier solves a strictly convex problem, whereas a L1 soft-margin classifier solves a convex problem. Convexity of the objective functions is one advantage of a SVM in comparison to other methods, e.g. ANNs which can have several local minimal. There is a trade-off between a large margin and the number of misclassifications. If the regularization parameter C is chosen small, the constraints can be more easily ignored and a large margin is obtained. On the other hand, if C is chosen large, the margin is small and less training samples are misclassified which can result in lower generalization. If $C \to \infty$, all constraints must be fulfilled which then equals a hard margin problem. Furthermore, samples on the border of the margin are called unbounded support vectors. The equality sign in Eq. 5.13 is valid for these. Misclassified samples or samples within the margin are bounded support vectors, thereby the support vector $\mathbf{x}_k$ is correctly classified if $\xi_k < 1$ and misclassified if $\xi_k \geq 1$. When all data points except the support vectors are deleted from the data set, the optimization still results in the same hyperplane.

If the data from a two-class problem is mapped to a sufficiently high-dimensional feature space, a linear hyperplane can always be found which perfectly separates the two classes.

Given any feature map $\boldsymbol{\phi} : \mathbb{R}^d \to \mathbb{R}^m$, the following minimization is solved

$$\begin{aligned} &\min_{\mathbf{w}\in\mathbb{R}^d, \xi_k \in \mathbb{R}^+} \|\mathbf{w}\|^2 + C\sum_{k=1}^{N} \xi_k^p \\ \text{subject to:} \quad &y_k \ (\mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_k) + w_0) \geq 1 - \xi_k \quad \text{for} \quad k = 1...N \ . \end{aligned} \tag{5.14}$$

In this case, the classifier $f = \mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}) + w_0$ is linear for $\boldsymbol{\phi}(\mathbf{x}) \in \mathbb{R}^m$, but nonlinear for $\mathbf{x} \in \mathbb{R}^d$. Mapping to a more suitable feature space $\mathbb{R}^m$ decreases violation of the constraints, i.e. misclassifications of training samples, and increases generalization, when the training set is hardly linearly separable in $\mathbb{R}^d$. Note, m must not necessarily be larger than d, as long as the separability is improved. Considering that the solution $\mathbf{w}$ can always be expressed as $\mathbf{w} = \sum_{j=1}^{N} \alpha_j \boldsymbol{\phi}(\mathbf{x}_j)$, the constraints become

$$y_k \ \left(\sum_{j=1}^{N} \alpha_j \boldsymbol{\phi}^T(\mathbf{x}_j)\boldsymbol{\phi}(\mathbf{x}_k) + w_0 \right) \geq 1 - \xi_k \quad \text{for} \quad k = 1...N \ , \tag{5.15}$$

where the scalar product can be substituted by the kernel $k(\mathbf{x}_j, \mathbf{x}_k) = \boldsymbol{\phi}^T(\mathbf{x}_j)\boldsymbol{\phi}(\mathbf{x}_k)$. Applying the kernel trick, the optimization problem to solve is

$$\begin{aligned} &\min_{\alpha_k\in\mathbb{R}, \xi_k \in \mathbb{R}^+} \sum_{j,l=1}^{N} \alpha_j \alpha_l k(\mathbf{x}_j, \mathbf{x}_l) \ + C\sum_{k=1}^{N} \xi_k^p \\ \text{subject to:} \quad &y_k \ \left(\sum_{j=1}^{N} \alpha_j k(\mathbf{x}_j, \mathbf{x}_k) + w_0 \right) \geq 1 - \xi_k \quad \text{for} \quad k = 1...N \ . \end{aligned} \tag{5.16}$$

Instead of first calculating the mapping $\boldsymbol{\phi}$ for all data points and then calculating the scalar products, the kernel approach is faster. In addition, it allows using kernels for which the corresponding mapping $\boldsymbol{\phi}$ exists but is unknown. For any function $k(\mathbf{x}, \mathbf{x}')$ which is positive semidefinite and symmetric exists according to the Hilbert-Schmidt theorem a mapping function $\boldsymbol{\phi}(\mathbf{x})$ such that $k(\mathbf{x}, \mathbf{x}') = \boldsymbol{\phi}^T(\mathbf{x})\boldsymbol{\phi}(\mathbf{x}')$. Common kernels are the polynomial kernel $k(\mathbf{x}, \mathbf{x}') = (\mathbf{x}\mathbf{x}')^D$ with degree D and the radial basis function kernel $k(\mathbf{x}, \mathbf{x}') = \exp(-\gamma\|\mathbf{x} - \mathbf{x}'\|^2)$.

Until now, only a 2-class problem has been discussed. Extension of the approach to multiple classes are the one-against-all and the one-against-one method. In the former case, decision hyperplanes for the samples of one class against the rest is trained for each class. In the latter case, $c(c-1)/2$ hyperplanes are constructed separating two classes at a time and the final class decision is made on a majority vote. The unclassifiable region for the one-against-one method is smaller than for the one-against-all method [2].

Advantages of SVMs especially over neural networks are that they maximize generalization, have no local minimal, and are robust to outliers. Still, the trade-off parameter C, the kernel, and its parameters need to be selected. Also, computation time increases with an increasing number of training samples, especially for finding the best parameter selection. Within this work, an L1 soft-margin SVM with a radial basis function as kernel ($C = 1.0, \gamma = 0.01$) is used for one-against-one multi-class classification [20] .

5.3 Results for the Munich Database

In this section, each of the aforementioned feature extraction techniques for marker-based gait analysis is presented and discussed in the context of affect recognition. Challenges originating from the problem task and the characteristics of the gait database are:

- How are the trajectories over time best modeled? Therefore, two approaches are compared. On one hand, statistical features of the trajectories are taken and on the other hand a combination of PCA and FT is applied.
- The number of samples is small in comparison to the overall number of available features. Which features are best to chose for avoiding the curse of dimensionality? Several dimension reduction techniques are compared for this purpose. The linear techniques PCA and LDA, the non-linear methods KPCA and GDA, and feature selection based on statistics are investigated in particular.
- The task of locomotion primarily influences the gait recordings. Despite that, is it possible to retrieve information about the emotional state of a walker? Results on human performance suggest that accuracy should be at least above chance level.
- If information about the emotional state can be retrieved from a walker, which emotion model fits best to the capability of gait for emotion expression? A categorical and a dimensional model for emotions are compared for this reason.
- Gait patterns are suited for identification. Thus, how strong does the individual walking style influence recognition performance?

5.3.1 Data Preprocessing

The length of the recordings in the Munich database varies between 3 and 7 steps. This variation is caused by different stride lengths within a fixed recording area. To achieve better comparability, a single stride is extracted from each recording. The trajectory of the marker affixed to the heel is similar among all participants and shows a characteristic minimum; therefore this marker is chosen for extraction of a single stride. Fig. 5.4 illustrates over time the z-coordinate of the marker affixed to the right heel. A minimum indicates that the right heel touches the ground and is followed by a nearly constant position of the heel during the stance phase of the right leg. During the swing phase of the right leg, the foot leaves the ground and the height of the heel increases. An algorithm based on detecting these minima calculates the starting point of all strides in each recording. The algorithm first selects time intervals, in which height of the marker at the heel is less than 30% of maximum height, then disregards intervals which last less than 125ms and finally calculates the minimum in each interval. The complete recording of one gait is cut down to a single stride for all marker positions, joint centres and joint angles. Following feature extraction uses the joint angles during one stride. In comparison to marker positions, joint angles are relative to the walker and not to world coordinates. Still, speed, velocity, and stride length need to be calculated from the trajectories of the marker on the right heel.

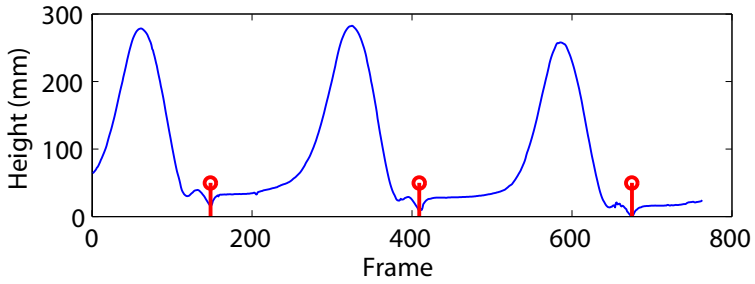


Fig. 5.4: The periodic characteristic of the marker affixed to the heel is one of the 105 marker trajectories. Here, a minimum, marked with a red bar, indicates the start of a stride.

Feature	NN	Naive Bayes	SVM
PCA-FT-PCA	43	41	57
Velocity, Cadence, Stride Length	52	45	45
Significant Subsection	**63**	49	47
Sig. Subsection + PCA (15PC)	58	52	**62**
Sig. Subsection + KPCA (15PC)	36	**60**	**60**
Sig. Subsection + LDA	**63**	55	**62**
All Joint Angles	56	45	25
All Joint Angles + PCA (30PC)	50	50	**69**
All Joint Angles + KPCA (23PC)	28	58	25
All Joint Angles + LDA	52	53	53

Tab. 5.3: Accuracy for inter-individual affect recognition in [%].

5.3.2 Inter-Individual Recognition

Concept of inter-individual recognition is that the sample set contains all recordings of the walker, who is left out in the training set. This is comparable to recognizing the affective state of an unknown walker. Iteratively, the accuracy is calculated for each walker left out. Table 5.3 compares accuracy for different feature extraction methods. If extra success comparing to a random predictor is above 45%, results are marked bold. Highest accuracy of 69% is achieved when PCA is applied to all joint angles and the first 30 principal components are used. Although accuracy is with 69% above chance level and comparable to human performance, which is 63% see chapter 3.3.1, this still means that one third of the samples are misclassified. It is concluded that without further knowledge of the walker such as identity, both individual differences in walking style and expression of affect complicate reliable estimation of affect purely based on the kinematics of walking.

Feature	NN	Naive Bayes	SVM
PCA-FT-PCA	70	70	78
Velocity, Cadence, Stride Length	85	83	76
Significant Subsection	87	**93**	**89**
Sig. Subsection + PCA (15PC)	**91**	85	**89**
Sig. Subsection + KPCA (15PC)	87	75	88
Sig. Subsection + LDA	**93**	**93**	**93**
All Joint Angles	**91**	**93**	79
All Joint Angles + PCA (15PC)	**92**	**92**	**95**
All Joint Angles + KPCA (15PC)	88	47	25
All Joint Angles + LDA	47	45	47

Tab. 5.4: Average accuracy for person-dependent affect recognition in [%].

5.3.3 Person-Dependent Recognition

The previous subsection shows that person-dependent recognition of discrete affective states is accomplishable above chance level, but can be improved if individuality is considered. For this reason, the classifiers are trained individually for person-dependent recognition i.e., the training set of each walker contains nine exemplars of neutral, happy, sad and angry walking. One exemplar of each affective state is left out iteratively, so that the training sets are balanced. Accuracy is calculated for each walker separately and mean accuracy over all walkers is reported in Tab. 5.4. Extra success above 85% is marked bold. Interpretation of Tab. 5.4 allows conclusions about the performance of different feature extraction methods and comparison of inter-individual versus person-dependent recognition. In the latter case, it is obvious that accuracy increases strongly for person-dependent recognition and reaches a maximum of 95% accuracy, which means 93% extra success comparing to a random predictor.

Comparing the accuracy achieved with different feature extractions reveals a number of interesting points:

a) Extracting eigenpostures and eigenwalkers from the data sets including all joint angles leads to 78% accuracy. For observations which include only a single stride, the performance of PCA to extract relevant temporal information from time series is less efficient than applying basic statistical parameters such as minimum, mean and maximum to the joint angles over time.

b) Recognition rate above 80% is already achieved if classification uses only the features velocity, cadence and stride length. In accordance with [32, 126, 131], velocity of the movement contains fundamental information about the affective state of a walker. The discrete affective states neutral and angry are best distinguishable. Although

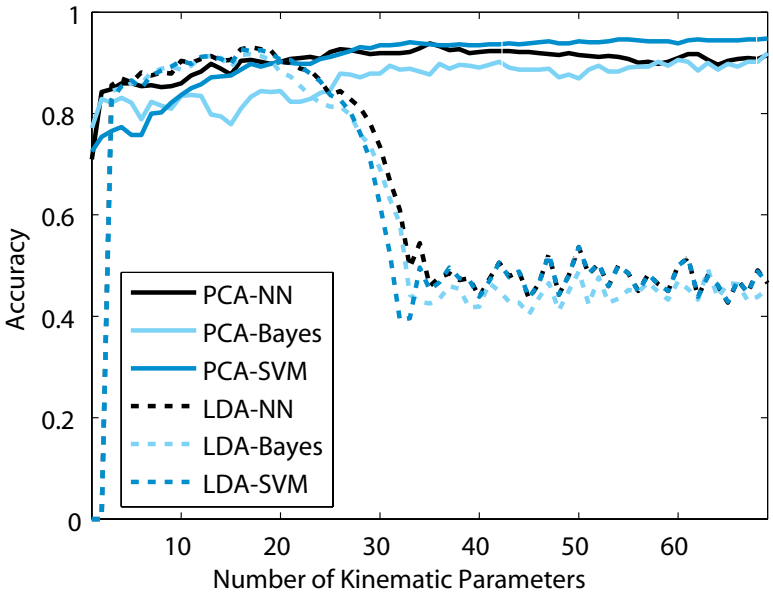


Fig. 5.5: Although LDA takes class affiliation into consideration, it outperforms PCA only for small numbers of kinematic parameters in relation to number of instances in the training set.

velocity gives already a good estimate on the affective state, estimation based only on this factor can be easily distorted in real-world scenarios.

c) Naive Bayes and NN perform well using the statistical parameters of all joint angles with 93% and 91%, respectively, even though the number of instances in the training sets is approximately half of the number of features. Reducing the number of features to 15 with PCA, leads to same performance of 92%-95% and thus does not discard relevant information regarding affect. Comparable accuracy of 87%-93% among all three classifiers is achieved when statistical parameters are calculated for only these joint angles which are referred in literature as significant [32, 131]. Therefore, one can conclude that reducing the number of statistical parameters by unsupervised PCA without any expert knowledge is as efficient for classification as involving expert knowledge, which has been obtained by statistical hypothesis testing beforehand.

d) As stated in literature, dimension reduction using LDA does not necessarily outperform PCA, though it considers class affiliations [99]. Fig. 5.5 illustrates the accuracy of PCA and LDA when the number of kinematic parameters is increased. The number of instances in the training set $n_{train} = 36$ is hold constant. PCA reduces the number of parameters to 15 PCs. For a small number of kinematic parameters, LDA reduces the feature space more efficiently than PCA. As soon as the within-class scattermatrix S_w becomes singular, which is the case if the number of kinematic parameters exceeds 32, accuracy of LDA stays around 40-50%. PCA is more robust against high dimensionality of the original space, if the original dimension exceeds the number of instances n_{train}. Accuracy of PCA even slightly increases by adding more kinematic parameters.

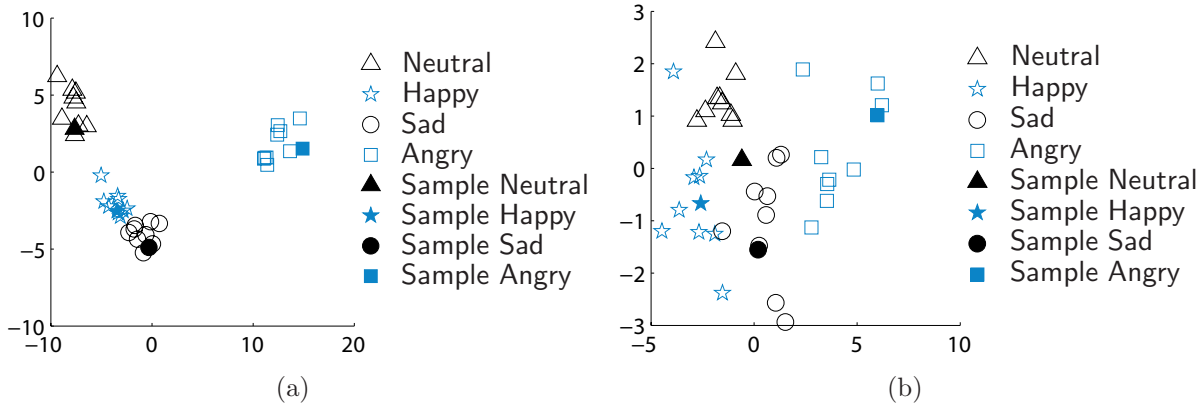


Fig. 5.6: Mapping the feature space on two dimensions with LDA graphically illustrates the difference in expressiveness among the walkers. (a) The cluster of angry gaits is clearly separable from the others for walker 7. (b) Clusters for each affective walking style of walker 4 are spatially less separable, so that misclassifications occur.

e) As noted in [151], nonlinear techniques do not necessarily outperform linear techniques for dimensionality reduction on real-world tasks. Here, KPCA extracts relevant features from the original parameter space, but overall performance is less than feature extraction with PCA. This result is in contrast to [156], who report an increase of 5% by using KPCA instead of PCA for a two-class separation task. So that one can conclude that the advantage of KPCA over PCA depends on the application. Results for GDA have not been added to Tab. 5.3 and 5.4, because recognition rates have not exceeded chance level.

f) Although Naive Bayes, SVM, and NN have different assumptions on the data, accuracy differs in general little among the classifiers, especially when feature space transformation such as PCA or LDA has been applied before. Low performance of SVM in some cases is traced back to the fact that the parameter γ of the kernel function and regularization parameter C have not been optimized for each particular case. Good performance of NN reveals that expression of different affective states forms separable clusters in the feature space. Naive Bayes assuming normally distributed features performs well in the person-dependent case, whereas a mixture of Gaussians for each feature caused by individuality of the walkers complicates inter-individual recognition with Naive Bayes.

Person-dependent recognition reveals the optimal accuracy which is achievable by excluding inter-individual differences in expressing affect and walking styles. Accuracy is only affected by a person's fluctuations in walking and acting a specific affective state. Fig. 5.6 graphically illustrates the differences among the walkers by mapping the feature space of the significant subset onto two dimension with LDA. One walker with high recognition accuracy and one with low accuracy are chosen for comparison. If LDA is applied to the 40 recordings of a single walker, separability of the clusters for each affective state depends only on the expressiveness of a walker and the size of the clusters depends on the variations in walking and expressiveness.

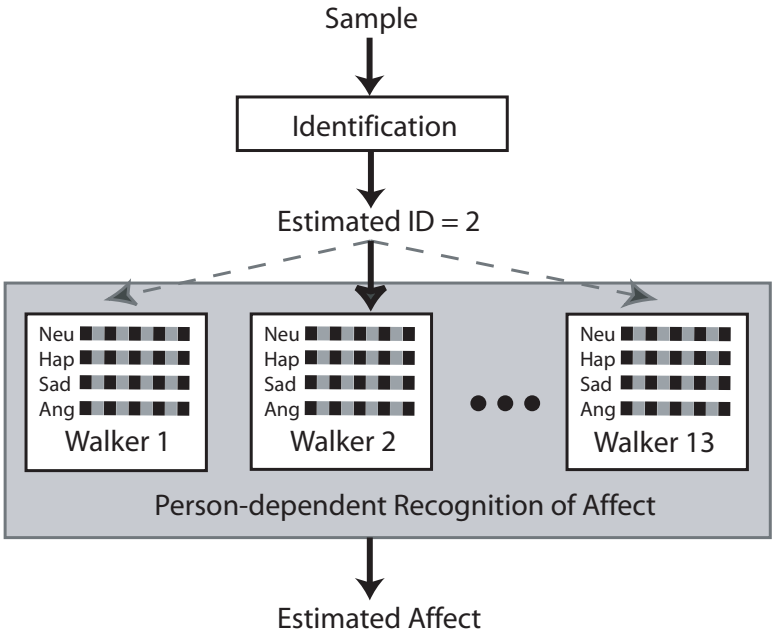


Fig. 5.7: In a two-stage classification, first the identity of a walker is estimated and then person-dependent recognition of affect is performed based upon the estimated identity.

Comparing person-dependent with inter-individual recognition advises to integrate identity in estimation of affective states. The next subsection investigates what performance is achievable if the identity is estimated beforehand and following recognition of affective states is based on the estimated identity.

Person-Dependent Recognition based on Estimated Identity

Good performance of person-dependent recognition faces the problem that identity is not necessarily given. This subsection combines identification based on kinematic parameters with following recognition of affect. The concept is illustrated in Fig. 5.7. As in the previous sections, mean accuracy over all walkers is calculated for evaluation of this concept. Iteratively, four affective gait patterns of each walker are left out from the training sets. As shown in the previous section, extracting temporal information with PCA is less suitable than calculation of statistical parameters. For this reason, the approach based on eigenpostures and eigenwalkers is excluded from further analysis. Same holds for KPCA regarding dimension reduction.

Table 5.5 shows the results for identification based on kinematic parameters. In contrast to recognition of affect, identification performs poorly based only on the parameters velocity, cadence, and stride length. Best result is achieved, if LDA is applied to the statistical parameters of all joint angles. Note, that the training set contains $n_{train} = 13 \cdot 9 \cdot 4$ instances for identification, so that n_{train} is much larger than the dimension of kinematic parameters in this case and the within-class scattermatrix does not become singular. Thus, good performance of LDA for identification does not contradict good performance of PCA for recognition of affect considering that the statistical parameters of all joint angles form the basis. LDA applied to the statistical parameters of all joint angles in combination with

Feature	NN	Naive Bayes	SVM
Velocity, Cadence, Stride Length	34	23	17
Significant Subsection	94	85	95
Significant Subsection + PCA	98	86	87
Significant Subsection + LDA	92	90	93
All Joint Angles	99	98	71
All Joint Angles + PCA	99	97	99
All Joint Angles + LDA	**99.6**	**99**	**99**

Tab. 5.5: Accuracy for identification in [%].

Feature	NN	Naive Bayes	SVM
Significant Subsection	84	**89**	86
Significant Subsection + PCA	**90**	81	87
Significant Subsection + LDA	87	88	88
All Joint Angles	**91**	**93**	79
All Joint Angles + PCA	**92**	**89**	**91**
All Joint Angles + LDA	43	42	43

Tab. 5.6: Accuracy of affect recognition based on estimated identity in [%].

NN is used to estimate identity for following recognition of affect.

Resulting recognition of affect based on estimated identity performs almost as good as based on real identity, i.e. 93% recognition rate is achieved. This is traced back to the fact that identification is performed with 99.6% accuracy. Hence, the dynamics of kinematic parameters contain information of identity as well as affect. Recognition of the affective states of a walker benefits from taking the information in the kinematics about the identity into account. Knowledge about the walker increases accuracy significantly compared to recognizing the affective state of an unknown walker, see Tab. 5.3.

The Relevance of Affect for Identification

The strong influence of identity on recognition of affect, raises the question if also identification can be affected by different affective states of a walker. For this purpose, identification of walkers expressing one affective state, e.g. happy, is performed based on a training set which contains only trials of another affective state, e.g. neutral. Results are shown in Tab. 5.7. Identification has been performed by applying LDA to the statistical parameters of all joint angles and using NN as classifier, see Sec. 5.3.3 on feature extraction for identification. If the same affective state underlies both training and testing set, identity is

	Test Set			
Training Set	**Neutral**	**Happy**	**Sad**	**Angry**
Neutral	100	95	87	87
Happy	99	100	85	93
Sad	92	77	100	54
Angry	91	86	86	100

Tab. 5.7: Identification under different affective states in [%].

estimated with 100% accuracy for the Munich database. However, if training and testing set differ in the affective state, which the walker expresses during walking, accuracy significantly decreases. Therefore, one can draw the conclusion that also identification suffers from interference caused by affect.

5.3.4 Recognition based on the PAD-Model

By analyzing the confusion matrix for recognition of discrete affective states in walking, the states sad and angry are generally better recognizable than the state happy. The states sad and angry differ highly in arousal, whereas arousal is similar for happy and angry, which differ mainly in pleasure. This leads to the assumption that the dimension arousal is better recognizable in walking than pleasure. As the dimensions pleasure, arousal, and dominance of the PAD-model are highly uncorrelated, recognition for each dimension is investigated separately. Person-dependent recognition distinguishes between low, medium, and high pleasure, arousal, or dominance. Accuracy is calculated for each walker separately using NN and mean accuracy for each dimension is listed in Tab. 5.8.

Regardless of feature extraction, mean accuracy for pleasure is lower than mean accuracy for dominance and arousal. Besides the expected high accuracy for arousal, also different levels of dominance are well distinguishable in gait patterns. Accuracy reaches 97% for arousal and 96% for dominance, when classification is applied to the statistical parameters of all joint angles. Accuracy for individual walkers is shown in Fig. 5.8 for this case. Different levels of pleasure are less recognizable than different levels of arousal and dominance for most walkers. Depending on the walker, either dominance or arousal is best distinguishable.

Looking only at the recognition performance using velocity, cadence, and stride length shows that different levels of arousal are with 91% accuracy distinguishable. Adding additional features increases more the separability for different levels of dominance and pleasure than for arousal. Thus by obseving only the three features velocity, cadence, and stride length, a good approximation on the activation of the walker can be given. However, relieble estimation of pleasure and dominance requires the observation of additional features.

The hypothesis, that different levels of pleasure are less recognizable in gait patterns for automatic recognition is confirmed. It is concluded that gait is more capable to reveal different levels of arousal and dominance than pleasure. This is also in accordance with

Feature	Pleasure	Arousal	Dominance
Velocity, Cadence, Stride Length	72	91	79
Significant Subsection	87	95	96
Sig. Subsection + PCA (7PC)	83	91	94
Sig. Subsection + LDA	80	91	92
All Joint Angles	88	**97**	**96**
All Joint Angles + PCA (7PC)	83	95	96
All Joint Angles + LDA	48	50	55

Tab. 5.8: Accuracy for affective dimensions in [%].

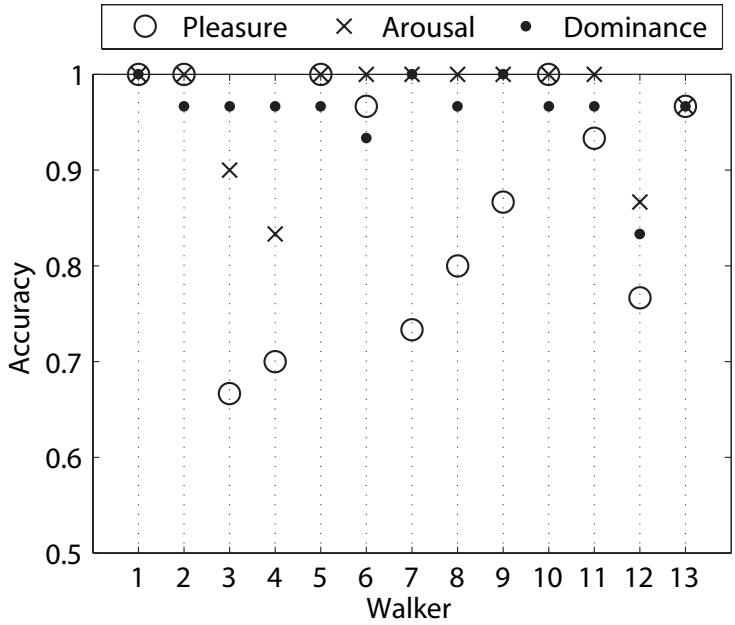


Fig. 5.8: Recognition rate of each walker for the dimension pleasure is lower than for the dimensions arousal and dominance. Depending on the walker different levels of pleasure are less recognizable than different levels of arousal or dominance.

human performance, see chapter 3.3.1.

5.3.5 Discussion

This section focuses on three central aspects regarding recognition of affective states in gait patterns, namely inter-individual versus person-dependent recognition, recognition based on discrete affective states versus extremes on the affective dimensions pleasure, arousal and dominance, and comparison of different feature extraction methods for marker-based gait analysis.

Although inter-individual recognition of affective states is accomplishable above chance

and in the range of human performance, person-dependent recognition outperforms the former one. Extra success comparing to a random predictor is twice as much. It is concluded that recognition is highly affected by individual walking styles and individual expression of affect. Identity and affect interact vice versa, so that accuracy for identification decreases, when classification is performed on affective samples which are not included in the training set.

Setting the results in comparison with published recognition rates for gesture and other motions, see Tab. 5.1, shows that recognition based on gait is more influenced by ones individual style than recognition based on gesture. Still, achieved inter-individual accuracy is larger than recognition based on the motion knocking [7]. Achieved person-dependent recognition rates are larger than reported accuracies of Bernhardt's *et al.s'* study [7] on knocking, Kapur's *et al.s'* study [73] on stylized body movements and Janssen's *et al.s'* study [69] on walking.

In accordance with related literature, which studies how humans perceive affect from gait patterns, automatic recognition also tends to recognize affective states which differ in arousal better than states which differ in pleasure [7, 126]. From this result it is concluded that gait is suited to deliver more easily cues about the activation of a walker.

Several approaches for feature extraction have been compared. Despite the fact that a combination of PCA and FT seems to be suitable to recognize the gender of a walker [150], it shows low performance for affect recognition. Calculation of statistical parameters from the time series results in higher accuracy. Even though the number of samples is approximately half the number of features for person-dependent recognition, classification without dimension reduction gives already good results for NN and Naive Bayes. The SVM mostly performs better after dimension reduction. Still reducing the dimension of the feature space to a number lower than the sample size, gives a more reliable and generalizable estimate on recognition accuracy. Depending on whether a significant subset or all joint angles are studied, either PCA or LDA performs better for dimension reduction. This is explained by a dependence on the number of training samples in comparison to the dimension of the feature space. It can be further noted, that even though the overall number of training samples is larger for inter-individual recognition, selection of relevant features is more crucial in this case than for person-dependent recognition. In addition, improvement in recognition performance by replacing linear with nonlinear techniques seems to be application-dependent in gait analysis.

Furthermore, automatic recognition uses a single stride in difference to human performance, which is based on videos lasting 7 seconds. As the captured gait kinematics have been visualized by an animated puppet, this is a potential hint that human perception from 2D motions is based on techniques which require observations containing at least three strides.

Due to its limitations and that people do not always walk, main application for recognition of affect based on gait is seen as an additional modality for multi-modal recognition of affect. It fills gaps in automatic recognition, when neither speech nor facial expression is available. Scenarios are high-security systems, human-robot-interaction, e.g. a human approaching a robot, and affective households.

As this study is based on acted affect, further studies are required for spontaneous

affect. This automatically requires video-based analysis, because a highly artificial setup for marker-based recording complicates repetitive elicitation of affect during walking. Further potential side aspects on recognition of affect from gait patterns are interference by gender, age, weight, or complaints of the walker.

The following challenges have to be accomplished for transferring recognition of affect in walking from laboratory to real-world. Still, marker-less video-based human motion reconstruction is not as accurate as marker-based systems [109]. Recognition of affect would definitely benefit from further refinements in retrieving detailed motion and thus kinematic parameters from video. This study shows that good recognition is already achieved using significant kinematic parameters of the upper body in combination with speed and speed-related parameters such as stride length and cadence. Hence, high accuracy in reconstruction of motion is especially required for the upper body. Furthermore, recognition is based on the kinematics of one side of the body. Due to the periodic and symmetric nature of gait, kinematics of the other side of the body, if available, can be used for increasing accuracy or verification of the measurements, outlier detection, or reconstruction of partially occluded areas. For integration in multi-modal emotion recognition systems, combination with previous action recognition, e.g. [86], is a prerequisite to deliver reliable recognition rates. Within this aspect, results of this study indicate that gait is suited for retrieving differences in arousal and dominance at distance.

5.4 Comparison on Emotion, Gender, Exhaustion, and Identity Recognition

The previous section investigates the recognition of emotions in gait patterns with several techniques from machine learning. Yet, not only the emotional state but also other factors such as age and gender influence a person's walking style. Two further marker-based gait databases have been recorded in the collaboration with the Institute of Biomechanics in Sports. One contains records of male and female walkers and the other has been recorded to investigate how exhaustion influences walking. Furthermore, the Emotive Motion Library, which is described in chapter 3.2.1, contains recordings of emotive walking styles. In the following, the common procedure for data analysis is described. Afterwards, the results are summarized for each of the three databases. It is noticeable that the NN classifier and the SVM calculate the same recognition rate for the all three databases if LDA is used for feature extraction. This aspect is analyzed in more detail, singularity of the matrix $\boldsymbol{S}_W$ is worked out as reason, and a mathematical proof is derived for the equivalence. Finally, the suitability to recognize emotion, gender, exhaustion, and identity from observing gait patterns is compared across the databases.

Two of the three databases provide reliable data only for marker positions and not for joint angles; therefore analysis is based on marker positions in this subsection. A common set of 23 markers is selected which contains the four markers at the headband, the markers on the vertebra C7, on the upper thorax, on the sternum, and on the left as well as right shoulder, on the elbow, on the interior and exterior wrist, on the ankle, toe, knee and heel. Minimum, mean, and maximum of each of the x-, y-, and z-trajectories are

calculated for each marker. This results in 9 features for each marker and, for 23 markers, in the input vector $\mathbf{x} \in \mathbb{R}^{207}$. In comparison to joint angles, marker positions are given in world coordinates. Thus, using marker positions as input for classification can bring – in particular for small databases – the drawback that the algorithms use information on the absolute position and walking direction for classification. To avoid this, each recorded data is set in reference to the marker on the vertebra C7. Furthermore, the data is transformed in the way that the walking direction is along the y-axis. In doing so, one feature which is the minimum of the y-direction of the marker C7, is always zero. Discarding this feature, leads to $\mathbf{x} \in \mathbb{R}^{206}$. It should be further noted that a single stride is extracted from each recording in the Exhaustion and Gender Database. Sequences of straight walking are short in the Emotive Motion Library. Highest common length of all trials is a single step. In this case, results are calculated on observation of a single step.

In the following, classification is applied either to all features or after dimension projection with PCA, LDA or KPCA. If results for KPCA are above chance level, they are included in the Tables 5.9, 5.10, and 5.11. Applying GDA to the features $\mathbf{x}$ does not lead to noteworthy results above chance level.

5.4.1 Exhaustion Database

Primary purpose for this gait database was to study the influence of exhaustion after a physical exercise. It was recorded by W. Seiberl at the Institute of Biomechanics in Sports, TU München. It contains the gait of 14 male subjects (mean age 25.3 ± 2.7 years; body mass index $23.5 \pm 1.9\text{kg/m}^2$) three times before and three times during exhaustion [173]. A program of warm-up and exercises at a rowing ergometer was performed before the recording of the exhausted condition. This database can be investigated with regard to identification and recognition of physical condition. As the recording quality for one subject is low, one data set was discarded from the database leaving 13 times three normal and 13 times three exhausted gait trials for classification. Results are summarized in Tab. 5.9.

Identification of the walkers performs with 100% accuracy (chance: 8%). Distinguishing between normal and exhausted walking style is only above chance (50%) if the training set contains gait trials of each walker. Recognition of the gait style of an unknown walker hardly exceeds chance level. Reasons are that the number of walkers is too less for inter-individual recognition, level of exhaustion differs among the walkers, and that expression of exhaustion is individual. KPCA achieves similar recognition rates if a polynomial kernel with $d = 3$ is used. After KPCA, performance of SVM is around chance level; therefore the parameter γ has been adjusted if required.

5.4.2 Gender Database

This database was also recorded by W. Seiberl at the Institute of Biomechanics in Sports, TU München. It contains the gait of 10 male subjects and 10 female subjects. A number of 10 repetitions of each walker are used for the data analysis.

[1] $\gamma = 1 \cdot 10^{-10}$ in this case.

	Feature	NN	Naive Bayes	SVM
Identity	All Markers	**100**	91	**100**
	All Markers + PCA(10PC)	96	99	69
	All Markers + KPCA(42PC)	98	54	95[1]
	All Markers + LDA(12EV)	82	69	82
Physical Condition	All Markers	83	65	53
	All Markers + PCA(30PC)	83	**87**	**87**
	All Markers + KPCA(47PC)	76	64	74[1]
	All Markers + LDA(1EV)	51	44	51
Physical Condition (inter-ind.)	All Markers	**60**	51	46
	All Markers + PCA(20PC)	49	**60**	53
	All Markers + LDA(1EV)	46	44	46

Tab. 5.9: Accuracy for the database on exhaustion in [%].

	Feature	NN	Naive Bayes	SVM
Identity	All Markers	97	98	95[2]
	All Markers + PCA(20PC)	**99**	96	98
	All Markers + KPCA(142PC)	90	64	79[3]
	All Markers + LDA(19EV)	97	88	97
Gender	All Markers	98	95	98[2]
	All Markers + PCA(30PC)	99	97	**100**
	All Markers + KPCA(137PC)	93	86	94[1]
	All Markers + LDA(1EV)	65	64	65

Tab. 5.10: Accuracy for the gender database in [%].

Calculated recognition rates are reported in Tab. 5.10. Accuracy for identification achieves 99% (chance: 5%). Also, recognition of gender is with 100% possible for this database. Performance of the SVM depends on γ; therefore this parameter has been adjusted in several cases as noted by the footnotes. Furthermore, a polynomial kernel with degree of five is best for KPCA in the case of gender recognition.

[2] $\gamma = 1 \cdot 10^{-5}$ in this case.
[3] $\gamma = 1 \cdot 10^{-17}$ in this case.

	Feature	NN	Naive Bayes	SVM
Identity	All Markers	47	50	42[4]
	All Markers + PCA(20PC)	57	46	57
	All Markers + KPCA(82PC)	48	17	46[6]
	All Markers + LDA(28EV)	**79**	52	**79**
Gender	All Markers	82	85	86[5]
	All Markers + PCA(25PC)	90	90	**95**
	All Markers + KPCA(30PC)	81	58	52
	All Markers + LDA(1EV)	66	66	66
Emotion	All Markers	40	50	25
	All Markers + PCA(20PC)	45	**59**	48
	All Markers + KPCA(60PC)	50	28	47[1]
	All Markers + LDA(3EV)	31	28	31

Tab. 5.11: Accuracy for the Emotive Motion Library in [%].

5.4.3 Emotive Motion Library

The Emotive Motion Library can be analyzed with regard to identify, gender, and emotion [98]. It provides one trial for each walker and for each of the four emotions. Thus, the training set contains only three instances of each walker for identification. Results for this database are summarized in Tab. 5.11. Recognition rate for identification is with 79% noticeable lower than for the previous databases. This is explained by a larger number of individuals and a lower number of training samples per class for the Emotive Motion Library. Still, good results are achieved for gender classification with 95%. Recognition rate for inter-individual emotion recognition is 59%. Performance of inter-individual recognition is comparable to the results for the Munich database, see subsection 5.3.2, even though the number of walkers is larger and the number of trials per walker is less. Person-dependent recognition rates can not be calculated because the database provides only one sample for each walker and each emotion.

5.4.4 LDA and the Small Sample Size Problem: The Impact of a Singular Within-Class Scatter Matrix on Classification

For all three databases, it has been observed that NN calculates exactly the same recognition rate than the SVM after LDA has been applied to reduce the dimension of the feature

[4] $\gamma = 1 \cdot 10^{-3}$ in this case.
[5] $\gamma = 1 \cdot 10^{-6}$ in this case.
[6] $\gamma = 1 \cdot 10^{-20}$ in this case.

space, see Tables 5.9, 5.10, and 5.11. This is regardless of the number of chosen eigenvectors that define the dimension of the feature space for the classification algorithm. Taking a closer look at the feature vector after LDA shows that the training samples of each class are mapped on a single point. This provides perfect separation of the training samples. However, test samples are not exactly mapped on this point which leads to an error rate for the recognition task. For this case, a mathematical derivation can be deduced. It shows that if the training samples of each class are mapped on a single point, the decision borders found by a maximum-margin classifier are the same as the decision border for nearest neighbor classification. This is observed for the small sample size problem where $\mathbf{S}_w$ is singular. The derivation first considers comparing the decision border of a NN and a hard-margin classifier for two classes. Then, the approach is generalized to SVMs and multiple classes.

If $\mathbf{S}_W$ is singular and the QZ algorithm is used to solve the general eigenvalue problem, it is observed for the three investigated databases that all samples of class $\mathcal{C}_i$ are mapped on a single point $\mathbf{p}_i$. Thus for a two-class problem, the vectors $\mathbf{p}_1 \in \mathbb{R}^m$ and $\mathbf{p}_2 \in \mathbb{R}^m$ are the representatives of each class after applying LDA to the training data. Considering that LDA aims to maximize the between-class separability while minimizing the within-class separability, mapping the samples for each class on a single point optimizes exactly this condition. This occurs in the case that the number of samples is small and $\boldsymbol{S}_W$ becomes singular.

The decision function f_{NN} of the NN algorithm is defined by

$$\begin{aligned} f_{NN}(\mathbf{x}) \geq 0 \quad \text{for} \quad y_i \quad &= 1 \quad (\mathbf{x}_i \in \mathcal{C}_1) \\ f_{NN}(\mathbf{x}) < 0 \quad \text{for} \quad y_i \quad &= -1 \quad (\mathbf{x}_i \in \mathcal{C}_2) \end{aligned}$$

where y_i is the class label for sample $\mathbf{x}_i$. The function f_{NN} separates the feature space in a way that a new sample is assigned to the class for which the Euclidean distance between sample and $\mathbf{p}_i$ is lowest. From a geometrical point of view, f_{NN} is a separation line with a normal vector pointing in the direction of $\mathbf{p}_1 - \mathbf{p}_2$. Thus, $f_{NN}(\mathbf{x})$ is:

$$f_{NN}(\mathbf{x}) = \frac{(\mathbf{p}_1 - \mathbf{p}_2)^T}{\|\mathbf{p}_1 - \mathbf{p}_2\|} \left[\mathbf{x} - \frac{1}{2}(\mathbf{p}_2 + \mathbf{p}_1)\right] \ . \tag{5.17}$$

In the above described case, the optimization problem for a hard-margin classifier is

$$Q(\mathbf{w}, w_0, \boldsymbol{\alpha}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{n_1+n_2} \alpha_i \{y_i(\mathbf{w}^T\mathbf{x}_i + w_0) - 1\} \ ,$$

where α_i are the non-negative Lagrange multipliers. As $\mathbf{p}_1$ and $\mathbf{p}_2$ are the only support vectors, the optimization problem can be reduced to

$$Q(\mathbf{w}, w_0, \boldsymbol{\alpha}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{2} \alpha_i \{y_i(\mathbf{w}^T\mathbf{p}_i + w_0) - 1\} \ . \tag{5.18}$$

Hence, the Karush-Kuhn-Tucker conditions are

$$\begin{aligned} \frac{\partial Q(\mathbf{w}, w_0, \boldsymbol{\alpha})}{\partial \mathbf{w}} &= 0 \ , & (5.19)\\ \frac{\partial Q(\mathbf{w}, w_0, \boldsymbol{\alpha})}{\partial w_0} &= 0 \ , & (5.20)\\ \alpha_i\{y_i(\mathbf{w}^T\mathbf{p}_i + w_0) - 1\} &= 0 \quad \text{and} \quad \alpha_i > 0 \quad \text{for} \quad i = 1, 2 \ . & (5.21) \end{aligned}$$

The following equations are derived from the the Karush-Kuhn-Tucker conditions:

$$\begin{aligned} \text{from} \quad 5.19 \ &: \ \mathbf{w} = \alpha_1\mathbf{p}_1 - \alpha_2\mathbf{p}_2 \ , & (5.22)\\ \text{from} \quad 5.20 \ &: \ \alpha_1 = \alpha_2 \ ,\\ \text{from} \quad 5.21 \ &: \ \mathbf{w}^T\mathbf{p}_1 + w_0 = 1 \ ,\\ &\quad\ -\mathbf{w}^T\mathbf{p}_2 - w_0 = 1 \ . \end{aligned}$$

Solving the equation for α_i gives $\alpha_i = \frac{2}{\|\mathbf{p}_1 - \mathbf{p}_2\|^2}$ and inserting in Eq. 5.22 provides $\mathbf{w} = \frac{2(\mathbf{p}_1 - \mathbf{p}_2)}{\|\mathbf{p}_1 - \mathbf{p}_2\|^2}$. With $w_0 = -\frac{(\mathbf{p}_1 - \mathbf{p}_2)^T(\mathbf{p}_1 + \mathbf{p}_2)}{\|\mathbf{p}_1 - \mathbf{p}_2\|^2}$, the decision function for a hard-margin classifier f_{HM} is

$$f_{HM}(\mathbf{x}) = 2\ \frac{(\mathbf{p}_1 - \mathbf{p}_2)^T}{\|\mathbf{p}_1 - \mathbf{p}_2\|^2}\left[\mathbf{x} - \frac{1}{2}(\mathbf{p}_1 + \mathbf{p}_2)\right] \ . \tag{5.23}$$

Comparing Eq. 5.17 with Eq. 5.23 leads to

$$f_{HM} = \frac{2}{\|\mathbf{p}_1 - \mathbf{p}_2\|} f_{NN} \ . \tag{5.24}$$

As class assignment relies only on the sign of the decision function, a hard-margin classifier achieves the same classification results as a nearest neighbor classifier. Considering that all samples of a class i are mapped on a single point $\mathbf{p}_i$, performance of a soft-margin classifier is the same as performance of a hard-margin classifier.

In the following the comparison is extended to SVM with a kernel $k(\mathbf{x}, \mathbf{x}')$. In this case, the dual problem to be solved can be written as

$$Q_{SVM,dual}(\boldsymbol{\alpha}) \ = \ \sum_{i=1}^{2} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{2} \alpha_i\alpha_j y_i y_j k(\mathbf{p}_i, \mathbf{p}_j) \tag{5.25}$$

and with $\sum_{i=1}^{2} y_i\alpha_i = 0$ follows $\alpha_1 = \alpha_2 = \alpha$ and

$$Q_{SVM,dual}(\alpha) = 2\alpha - \frac{1}{2}\alpha^2 \sum_{i,j=1}^{2} y_i y_j k(\mathbf{p}_i, \mathbf{p}_j) \quad \text{subject to} \quad \alpha > 0 \ . \tag{5.26}$$

For a soft-margin classifier, α should be bounded by an upper level. However, this can be discarded for the following calculations, because a soft-margin classifier calculates the same decision border as a hard-margin classifier in the studied case as mentioned above.

Setting the derivative $\frac{dQ_{SVM}(\alpha)}{d\alpha}$ equal to zero, leads to

$$\alpha = \frac{2}{k(\mathbf{p}_1, \mathbf{p}_1) - 2k(\mathbf{p}_1, \mathbf{p}_2) + k(\mathbf{p}_2, \mathbf{p}_2)} \ .$$

Thus, the decision function is

$$f_{SVM}(\mathbf{x}) = \frac{2}{k(\mathbf{p}_1, \mathbf{p}_1) - 2k(\mathbf{p}_1, \mathbf{p}_2) + k(\mathbf{p}_2, \mathbf{p}_2)} \sum_{i=1}^{2} y_i k(\mathbf{p}_i, \mathbf{x}) + b \ .$$

The value b is calculated by considering that for the unbounded support vectors j the equality $y_j = \ \alpha \sum_{i=1}^{2} y_i k(\mathbf{p}_i, \mathbf{x}) + \ b$ is valid

$$b = \frac{-\alpha}{2}[k(\mathbf{p}_1, \mathbf{p}_1) - k(\mathbf{p}_2, \mathbf{p}_2)] \ .$$

Hence, the decision function of a SVM for the studied case is:

$$\begin{aligned} &f_{SVM}(\mathbf{x}) = \\ &= \frac{2}{k(\mathbf{p}_1, \mathbf{p}_1) - 2k(\mathbf{p}_1, \mathbf{p}_2) + k(\mathbf{p}_2, \mathbf{p}_2)} \left\{ \sum_{i=1}^{2} y_i k(\mathbf{p}_i, \mathbf{x}) - \frac{1}{2} \left[k(\mathbf{p}_1, \mathbf{p}_1) - k(\mathbf{p}_2, \mathbf{p}_2) \right] \right\} \end{aligned} \quad (5.27)$$

Depending on the chosen kernel, equal class assignment can be derived for SVM and NN. If the kernel is a Gaussian $k(\mathbf{x}, \mathbf{x}') = \exp(-\gamma)\|\mathbf{x} - \mathbf{x}'\|^2$, the decision function $f_{SVM,g}$ is:

$$f_{SVM,g}(\mathbf{x}) = \frac{1}{1 - \exp(-\gamma\|\mathbf{p}_1 - \mathbf{p}_2\|^2)} \left\{ \exp(-\gamma\|\mathbf{p}_1 - \mathbf{x}\|^2) - \exp(-\gamma\|\mathbf{p}_2 - \mathbf{x}\|^2) \right\} \ . \quad (5.28)$$

As $\exp(-\gamma\|\mathbf{p}_1 - \mathbf{p}_2\|^2)$ is always lower than one, the term $(1 - \exp(-\gamma\|\mathbf{p}_1 - \mathbf{p}_2\|^2))^{-1}$ can be seen as a scaling factor. Thus, class assignment depends only on the sign of $\{\exp(-\gamma\|\mathbf{p}_1 - \mathbf{x}\|^2) - \exp(-\gamma\|\mathbf{p}_2 - \mathbf{x}\|^2)\}$. Considering that class assignment for the NN classifier depends also only on the sign of $(\mathbf{p}_1 - \mathbf{p}_2)^T[\mathbf{x} - \frac{1}{2}(\mathbf{p}_2 + \mathbf{p}_1)]$, see Eq. 5.17, the absolute value of $f_{SVM,g}(\mathbf{x})$ and $f_{NN}(\mathbf{x})$ is not relevant for comparison. Instead, it is sufficient that the sign of $f_{SVM,g}(\mathbf{x})$ and $f_{NN}(\mathbf{x})$ is the same for equal classification results. To assign the label $y = 1$, it is required for the SVM that $f_{SVM,g}(\mathbf{x}) > 0$ and consequently

$$\exp(-\gamma\|\mathbf{p}_1 - \mathbf{x}\|^2) > \exp(-\gamma\|\mathbf{p}_2 - \mathbf{x}\|^2) \ . \quad (5.29)$$

To get the same class assignment with a NN classifier, the following inequality

$$\mathbf{p}_1^T(\mathbf{x} - \frac{1}{2}\mathbf{p}_1) > \mathbf{p}_2^T(\mathbf{x} - \frac{1}{2}\mathbf{p}_2) \quad (5.30)$$

must be fulfilled. By showing, that if the inequality 5.29 is true also the inequality 5.30 is true, it is deduced that under the studied case the class assignment of a NN classifier equals class assignment of a SVM with a Gaussian kernel for all $\mathbf{x} \in \mathbb{R}^m$. The required

steps starting from inequality 5.29 are:

$$\begin{aligned} \|\mathbf{p}_1 - \mathbf{x}\|^2 &< \|\mathbf{p}_2 - \mathbf{x}\|^2 \\ \mathbf{p}_1^T\mathbf{p}_1 - 2\mathbf{p}_1^T\mathbf{x} + \mathbf{x}^T\mathbf{x} &< \mathbf{p}_2^T\mathbf{p}_2 - 2\mathbf{p}_2^T\mathbf{x} + \mathbf{x}^T\mathbf{x} \\ \mathbf{p}_1^T(\mathbf{p}_1 - 2\mathbf{x}) &< \mathbf{p}_2^T(\mathbf{p}_2 - 2\mathbf{x}) \\ \mathbf{p}_1^T(\mathbf{x} - \frac{1}{2}\mathbf{p}_1) &> \mathbf{p}_2^T(\mathbf{x} - \frac{1}{2}\mathbf{p}_2)\ . \end{aligned}$$

Thus, for a two-class problem, a SVM with a Gaussian kernel, a hard margin classifier and a simple NN classifier achieve exactly the same recognition rate if LDA is applied to a data set and the problem of a singular $\boldsymbol{S}_W$ is solved by using the QZ algorithm to solve the eigenvalue problem for LDA.

The number of decision borders increases for a multi-class problem with $c > 2$. The traditional SVM is defined for two classes and the one-against-one method is selected for extension to multiple classes. Considering only a pair of two classes out of the c classes, the decision border for this pair is the same when NN, a hard margin classifier or an SVM with a Gaussian kernel is used for classification. Hence, all the decision borders between two out of the c classes are equal for the three classification techniques and the three techniques assign the same class label to an unknown sample.

It should be further noted for the studied case that all training samples of a class are mapped by LDA on a single point $\boldsymbol{p}_i$. But unknown samples are not necessarily mapped on these points $\boldsymbol{p}_i$. Hence, if neither leave-one-out or cross-validation is applied in the studied case, classification rate for the three classifiers would be 100%. However, this rate would give no useful information.

The following conclusions are drawn from this derivation. When LDA is applied to a small sample size problem, the within-class scatter matrix $\boldsymbol{S}_W$ becomes singular. Still, the resulting eigenvalue problem can be solved with the QZ algorithm. Then, all training samples of each class i are mapped on a single point $\boldsymbol{p}_i$. In doing so, the between-class separability is maximized while the within-class separability is minimal. For this case, it can be shown that a simple NN classifier calculates the same decision borders as a hard-margin classifier or a SVM with a Gaussian kernel. Hence, the same recognition rates are calculated by the techniques. Even though applying LDA to a small sample data set seems to be not optimal, still comparable results to other techniques can be achieved, see Tab. 5.9, 5.10 and 5.11. Also, Deng *et al.* report good results for LDA using the QZ algorithm even if $\boldsymbol{S}_W$ is singular for the small sample size case [39].

The null space method has been proposed if $\boldsymbol{S}_W$ becomes singular for LDA [24, 64]. The derived equivalence of NN, hard margin classifier, and SVM with a Gaussian kernel is also valid if this technique is applied to the small sample set problem. If in these cases, one of the classifier needs to be chosen, the decision can be made with less focus on recognition rate and more focus on other aspects such as computational complexity. Also for comparing different feature extraction techniques, it is sufficient to apply only one of the above mentioned classifiers to the data set. Applying the other classifiers would give no additional information. This is especially relevant the larger the feature space gets.

Database	Category	# Classes	# Test Samples	# Training Samples	Feature Dimension
Exhaustion	Identity	13	1×13	5×13	206
	Physical Condition	2	13×2	26×2	206
	· Inter-individual	2	3×2	36×2	206
Gender	Identity	20	1×20	9×20	206
	Gender	2	10×2	90×2	206
Emotive Motion Library	Identity	29	1×29	3×29	203
	Gender	2	14, 15	42, 45	203
	Emotion	4	1×4	28×4	203
Munich Database	Identity	13	1×13	39×13	69
	Emotion[7]	4	1×4	9×4	69
	· Inter-individual	4	10×4	120×4	69

Tab. 5.12: Characteristics of the databases.

5.4.5 Comparison and Discussion

Comparison of the different databases barely on the recognition rate is not sufficient. Instead interpretation of the recognition rate needs to consider different training sizes, sample sizes, numbers of classes, and dimensions of the feature vector. Tab. 5.12 summarizes the characteristics of the databases. The number of training and test samples are given for each iteration cycle of the leave-one-out procedure. The total number of samples for each database is 78 for the exhaustion database, 200 for the gender database, 116 for the Emotive Motion Library, and 520 for the Munich database. The Munich database is investigated in subsection 5.3 using joint angles as input. If marker positions are taken instead of joint angles, recognition rates slighlty vary from the reported results. As the variations are only small, the results from the previous section can be used for the following comparison for better consistency with the previous section.

Fig. 5.9 plots the accuracy for each database over the number of classes. The chance level is in inverse ratio to the number of classes. Identification performs with a high recognition rate for all databases despite a large number of classes. Person-dependent recognition rate is similar to the recognition rate of gender. Both factors have a small number of classes in common. Thus, comparing the recognition of gender or the emotional state of a kown walker with recognition of a person leads to the conclusion that distinguishing a person from a limited group of walkers is best achievable. Lower accuracy for identification of the Emotive Motion Libary is explained by the fact that the training set contains only three trials of each walker which differ in the expressed affect. Trying to recognize the physical condition or the affective state of an unknwon walkler is about chance. The according

[7] Final recognition rate is averaged over the 13 walkers.

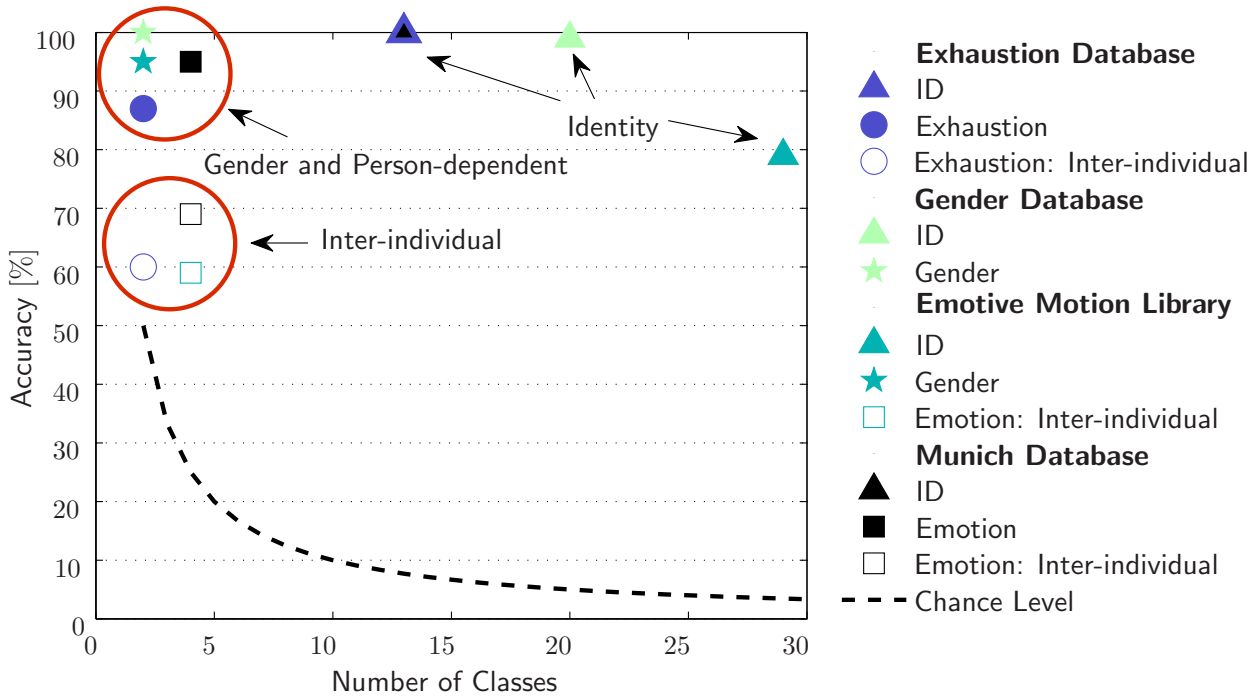


Fig. 5.9: Recognition performance varies depending on the factor. Best rates are achieved for identification, depicted by triangles. Independent of the database, inter-individual recognition of physical or affective state form a distinct cluster which is clearly below the cluster for person-dependent recognition. Person-dependent recognition is in a similar range as gender recognition. The dashed line depicts the chance level.

rates for inter-individual recognition form a distinct separable cluster which lies clearly below person-dependent recognition. It should be noted, that the inter-individual rates of the Munich database and the Emotive Motion Library for emotion recognition lie in the same range even though they differ in the number of training samples and in the recording procedure. Recording of the Emotive Motion Library was based on telling all participants a common story for each affective state. For the Munich Database, each walker memorized a situation in which he had felt each affect.

Chapter 3.3 summarizes human performance to recognize a person, gender, or affective states from observing the gait. Performance of automatic classification lies in the range of human performance for emotion recognition. The investigated methods from machine learning are considerably better than human performance for gender recognition and identification. A possible reason is that the measures provided by the optical motion tracking system are more accurate than the accuracy of the human eye for measuring e.g. the shoulder or hip width at distance. Both gender recognition and identification rely on observing such static features which describe also the physique.

5.5 Summary

To provide a means for emotion recognition at distance, this chapter studies the recognition of emotions in gait patterns with techniques from machine learning. Focus lies on the comparison of several pattern recognition algorithms, of two emotion models for recognition, of person-dependent versus inter-individual recognition, and of emotion recognition versus gender and identity recognition. The results and contributions are shortly summarized in the following.

Pattern recognition is divided in data preprocessing, feature extraction, and classification. At first, a single stride is extracted from each data set so that reported recognition rates refer to the observation of a single stride. As the recorded gait database contains highly dimensional, temporal dependent, and highly variable data vectors, efficient feature extraction is necessary. Therefore, a number of techniques is compared which are feature selection, PCA, LDA, and the according kernel extensions. Depending on the number of training samples and classes, either PCA or LDA performs better. The nonlinear techniques GDA and KPCA do not reach higher recognition rates than the linear techniques. Even though the classifiers NN, SVM, and Naive Bayes have different assumptions on the data, the recognition rates are similar. For extracting the temporal information in the gait trajectories, a combination of PCA and FT has been investigated. Yet, simple calculation of minimum, mean, and maximum of each trajectory achieves higher recognition rates.

The categorical versus the dimensional emotion model are compared for their suitability to recognize affect in gait. A set of categorical emotions is recognizable, yet the dimensional model better reveals that in particular differences in arousal and dominance are more accurately recognized than differences in pleasure.

The human gait is a highly individual motion pattern. Therefore, algorithms reach higher accuracy if they are trained for an individual. Practical applications benefit if first the identity is estimated and then the emotion is recognized. Likewise, it is shown that emotions expressed during walking can decrease the quality of identification if the training database contains only neutral walking trials.

Comparing the recognition of emotions with the recognition of gender and identity shows that person-dependent emotion recognition is in the range of gender recognition. However, accuracy for inter-individual emotion recognition is less than for gender recognition or identification.

If LDA is applied to reduce the dimensionality of the feature space and only a small number of training samples are available, the within-class scatter matrix $\boldsymbol{S}_W$ becomes singular. Using the null space method or the QZ algorithm to solve the eigenvalue problem leads to equal recognition rates for NN and SVM. A general proof is derived that NN, hard margin classifier, and SVM with a Gaussian kernel define the same decision borders for this case.

In general, the rates to recognize emotions in gait are similar to human performance. Even though the gait is a highly individual motion pattern, recognition rates are above chance. Hence, observing the gait can be considered to recognize the affective state at distance for a group consisting of a limited number of known persons.

5.6 Limitations

This approach shows that emotions are recognizable with techniques from machine learning, but the study is also accompanied by a number of limitations.

First, the investigated gait databases contain acted and elicited emotion. No gait database currently exists which contains spontaneous emotions. Still, the presented results provide an upper estimate on the recognition rates for spontaneous emotions. Recording spontaneous emotions would require a natural setup. This directly demands for highly accurate computer vision algorithms which extract the silhouette of the walker and model his/her pose.

Second, the human gait is influenced by many factors such as body pose, age, weight, physique, complaints, and gender. Within this study, the identity of the walker has been considered. A future study may also investigate to what extend emotion recognition is affected by variations in the other factors.

Third, recognition rates for emotions in gait are achieved above chance level. Yet, these rates are lower than comparable recognition rates for facial expressions and speech, in particular for differences in pleasure. The gait seems to be more suited to provide information about arousal or dominance of a walker.

Lastly, the term emotion is extensive. It could be also considered if gait is more suited to monitor long-term states such as mood. A possible application would be in therapeutic support for depression, which significantly affects the walking style, e.g. in the framework of embodied emotions.

Summing up, future work may contain to develop computer vision algorithms for recognition of emotions, to record a database which contains spontaneous emotions, and to investigate the impact of other factors on the gait.

6 HMM for Recognition of Affect and Identity in Marker-Based Gait Analysis

The previous chapter 5 focuses on static classifiers for recognition of affect in gait patterns. As walking is a process evolving over time, this chapter considers a classifier which is especially suited to handle sequential data. Within this context, a Hidden Markov Model (HMM) is most popular for sequential data in pattern recognition. Characteristic for an HMM is that the states of the system remain hidden, and the current state is deduced from the observations. HMMs have been introduced for identification in vision-based gait analysis [10, 71, 139]. The principle is that a stride is divided in several stances and the transition from one stance to the next is modeled by a Markov Chain. Thus, the current stance is a hidden state which can be inferred from the observations. An HMM is trained for each class which is either the identity or the affective state within this chapter. The probability which model emits most probably the observation sequence is calculated during the recognition process.

HMMs have been applied to recognize the identity, age, or gender in vision-based gait analysis. Motivated by reported results, this concept is applied to marker-based gait analysis in the following. Three different approaches are implemented which are 1) a simple minimum distance classifier which considers different stances but does not model the transition between the stances, 2) an HMM which uses a set of joint angles as observation vector, and 3) an HMM which uses a distance metric as observation vector and has been applied in vision-based gait analysis. Achieved recognition rates of these approaches are compared with each other. Afterwards, they are set in comparison with the recognition rates for static classification which is investigated in chapter 5.

The purpose of this chapter is to study if the achieved recognition rates of chapter 5 can be further improved by using an HMM instead of static classification. Application of HMMs to vision-based gait analysis has been studied in several publications [10, 71, 139, 159, 162]. Yet, if its performance is superior to static classifiers with efficient feature extraction in the case of marker-based gait analysis and in the context of affect recognition has not yet been studied. To facilitate comparison of the recognition rates, this chapter uses the same joint angles, the same length of the observation sequences, which is a single stride, and the same evaluation procedure as in chapter 5. It should be noted that to shorten the observation sequence to a single stride is not a prerequisite for the use of an HMM because the underlying Markov Chain is especially suited to deal with observations differing in length.

This chapter starts with an overview on the use of HMMs in vision-based gait analysis and motivates its application for marker-based analysis in section 6.1. Then, the basic concept of an HMM, different models and implementation issues are briefly summarized in section 6.2. Results of the minimum distance classifier are given in section 6.3. The design and training algorithms of the applied HMMs are described, followed by a summary

on the recognition rates of the HMM approaches. Finally, the conclusion compares the presented approaches with static classification in terms of recognition rates in section 6.4 and limitations are discussed in 6.5.

6.1 HMM in Vision-based and Marker-based Gait Analysis

The predominant research in vision-based gait analysis focuses on the recognition of identity besides gender and age [10, 139, 159, 162]. Similar marker-based approaches can also be found in literature on classification of human movements such as running, walking, standing, etc. [86]. The advantage of gait in comparison to other biometrics such as fingerprint or signature is that it is non-obtrusive and gives an estimate from distance. Promising recognition rates are reported in [10, 71, 139] for vision-based gait analysis. Yet, it is seen rather as one component of a multimodal biometric system. The reason therefor is that the walking style of a person varies depending on the walking surface, footwear, clothes, and walking speed. This variation over time complicates applications outside laboratory setups and is a challenge for mathematical algorithms. Regarding ethical issues, preserving a person's privacy is an issue just as for other biometrics.

In the following, the methodology in vision-based gait analysis is briefly summarized. First, a human's silhouette is extracted from the frames. As in [139], this can be achieved by semiautomatical location of bounding boxes. Within each bounding box, the silhouette is extracted using fore- and background detection. This task can be facilitated for indoor recordings with a static and uniformly colored background. For more natural databases, other issues such as shadow detection, variations in lightning, and moving background play a role. These and related issues are addressed in [139, 140]. Second, single gait cycles are extracted from the gait sequences. Finally, different feature extraction and classification algorithms are applied to the data. Feature extraction is generally subdivided in model-based and appearance-based approaches. Feature-based or holistic are synonyms for the latter one in literature. Boulgouris *et al.* summarize advantages and disadvantages of several model-based and appearance-based features in [10]. Extracting model-based features requires a high quality of the sequences. However, it has the advantage that the features are view and scale invariant. Appearance-based features are derived from the silhouette of the walker and can handle a lower recording quality. Features fast to compute are the width of silhouette and its vertical and horizontal projection [10, 71]. Often better results are achieved if the silhouette is from the side view and not from the front view of the walker. Classification is either achieved by template-matching or an HMM. Due to different walking speeds, template matching is improved if time normalization is taken into account. Proposed algorithms within this context are Dynamic Time Warping and Linear Time Normalization [10]. The distance metric, upon which the test decision is made, is based on a comparison of a complete walking cycle between test sequence and reference sequence which is a results of the training. It does not take into account different stances during one cycle. An HMM subdivides a complete gait cycle in several stances and models the output during each stance. Classification results are often reported in Cumulative Match Characteristic (CMC) curves, in which the recognition rate is plotted over the rank. A rank of 5 denotes that the correct identity is among the first 5 reference

sequences which are closest to the test sequence.

In clinical and biomechanical studies, marker-based gait analysis is preferred over video-based because of a higher accuracy in estimation of model-based parameters. The data is either statistically analyzed or with several techniques from machine learning, see chapter 5.1.1. In marker-based gait analysis, HMMs are used to discriminate between different movements, e.g. walking, dancing, kicking [86]. The obtained models can further generate artificial motion primitives for computer animations or robots. Motivated from results in vision-based gait analysis and for motion discrimination, the HMM concept is adopted for marker-based gait analysis in the following. That way, single stances of a gait cycle are modeled, whereas in the previous chapter features are derived from a complete gait cycle. In doing so, the feature space needs to be reduced to enable convergence of the training algorithms for the HMM. To compare the efficiency of the algorithms between marker-based and vision-based gait recordings, recognition rates for identification are calculated. In the context of affective computing, the data is also classified regarding emotional states. Focus lies especially on the research question whether recognition rates can be further improved by the use of an HMM.

6.2 Hidden Markov Model

Classification can be divided in parametric and non-parametric techniques. Parametric techniques rely on a model of the process. If the model is known or well-estimated from the training data, parametric techniques can have the advantage to be optimal. Disadvantages of parametric techniques are that the exact model of a stochastic process is seldom known. Reasons therefore are high complexity of the system, limited training data, and limited knowledge on the underlying mechanism. As an HMM belongs to the class of parametric techniques, the latter issues are also crucial in training an HMM. In this section, an HMM is applied to the recorded gait patterns, which models the development over time of the gait trajectories.

The key element of a HMM is a finite set of discrete states $\omega_1, \omega_2, ..., \omega_I$. At time t, the system is in state $\omega(t)$. A sequence of a maximum number of T states is denoted by the state vector $\boldsymbol{\omega} = [\omega(1), \omega(2), ..., \omega(T)]$. If these states, also named latent variables, are observable, the system is called a Markov Chain. However, in most real-world applications, the states are hidden and not directly measurable. Then the observation $v(t)$ is made with a certain probability, if the system is in state $\omega(t)$. Thus, the underlying state is inferred from the observations, but ω itself remains hidden. Fig. 6.1 illustrates the concept both for the Markov Chain and for the HMM. Characteristic of an HMM is a discrete number of states while the observations can be either discrete or continuous. If also the state is continuous and can be modeled by a Gaussian, a linear dynamical system can be used to model the process. The following general description of HMMs is based upon the textbooks of Rabiner and Juang [129], Bishop [9], and Duda [43].

The assumption of a 1st order Markov Chain is that the Markov property is met. This is equivalent to causality where the future state $\omega(t+1)$ depends only on the current state

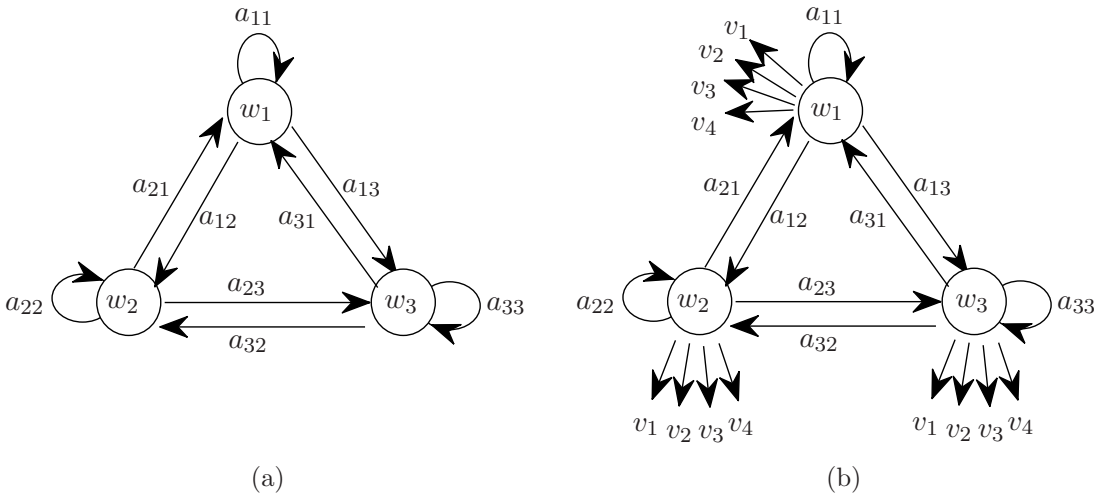


Fig. 6.1: (a) Transition between the states ω_1, ω_2, and ω_3 depends on the probabilities a_{ij} in a Markov Chain. (b) Only the observations v_j are visible in a Hidden Markov Model. The states ω_j remain hidden.

$\omega(t)$ and is independent of the past

$$P[\omega(t+1) = \omega_j | \omega(t) = \omega_i, \omega(t-1) = \omega_k, ...] = P[\omega(t+1) = \omega_j | \omega(t) = \omega_i] \ .$$

If $P[\omega(t+1) = \omega_j | \omega(t) = \omega_i]$ is time-independent, the matrix $\mathbf{A}$ contains the transition probabilities a_{ij} with $a_{ij} \geq 0$ and $\sum_j a_{ij} = 1$. If a diagonal entry of $\mathbf{A}$ is equal to one, the according state is called final or absorbing state. If all entries of $\mathbf{A}$ are unequal zero $a_{ij} \neq 0 \ \forall i, j$, the model is called ergodic, see Fig. 6.1. Yet, if all entries below the diagonal are equal zero, the model is a typical left-to-right model.

In an HMM, the state $\omega(t)$ is not measurable. A second random process models the probability of an observation $v(t)$ if the system is in state $\omega(t)$. A number of d observations are possible for discrete observations. The observation probability matrix $\mathbf{B}$ contains the probabilities $b_{jk} = P[v_k(t)|\omega_j(t)]$ to observe an observation v_k if the state is ω_j. The matrix $\mathbf{A}$ is a time-independent square matrix of dimension I, whereas $\mathbf{B} \in \mathbb{R}^{I \times d}$. A complete specification for a discrete density HMM (DDHMM) contains the set of model parameters $\Theta = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$, where $\boldsymbol{\pi}$ is the initial state distribution $[P(\omega(1) = \omega_1), P(\omega(1) = \omega_2), ..., P(\omega(1) = \omega_I)]^T$.

Three basic problems can be derived from the definition of an HMM:

- **Learning** Given the number of states I, the number of observations d, a structure of the model, and a set of training observations, how to estimate the probability matrices $\mathbf{A}$ and $\mathbf{B}$ of the model? The solution is the Baum-Welch algorithm, which is equivalent to the Expectation-Maximization (EM) method. It can result in local maxima and the results depend on the initial estimates of $\mathbf{A}$ and $\mathbf{B}$.

- **Evaluation** Given an HMM with $\Theta = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$, what is the conditional probability $P(\mathbf{V}|\Theta)$ that a sequence of observations $\mathbf{V} = [v(1), v(2), ..., v(T)]$ is generated by the model Θ? Straight forward calculation of $P(\mathbf{V}|\Theta)$ would be too computationally

intensive. Therefore the common solution is the forward or backward algorithm.

- **Decoding** Given a HMM with $\Theta = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$ and a set of observations $\mathbf{V}$, what is the most likely sequence of hidden states ω? The solution is the Viterbi algorithm.

A detailed mathematical description of the algorithms for a DDHMM is given in [9, 43, 128, 129].

Even though the discrete observations simplify the algorithms, this property limits the applicability of a DDHMM. One possibility is to apply vector quantization to convert continuous observations into a sequence of discrete observations. The other possibility is to use a continuous density HMM (CDHMM), where the observations are modeled by a probability density function which could be a linear combination of Gaussians. In the case of a single state $I = 1$, it is equivalent to a Gaussian Mixture model. Even though CDHMMs are more accurate than DDHMMs, they require more complex algorithms and a larger data base for training [26]. Further advanced HMMs exist which improve the limitations of a DDHMM. For example, the state transition probability does not only depend on the last state but also on several states in the past in a semi Markov Model. A factorial HMM contains multiple Markov chains. Then, the observation is a combination of the output of each Markov chain. This can decrease the required number of states for each chain, but requires additional complexity during training.

As HMMs include time-dependencies, they are a powerful technique. Still, the following issues need to be considered when training an HMM:

- **Insufficient Training Data:** Too short training sequences can lead to an inadequate number of occurrences of low-probability events. In the worst case, entries in the probabilities matrices are estimated to be zero, even though the true probability is only small. This would lead to misclassification of low-probability events. If the number of training samples can not be increased further, the number of model parameters $\Theta = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$ to be trained can be decreased. Another possibility is to define thresholds for the model parameters.

- **Initial estimates of the parameters of the HMM:** Choosing the initial estimates of $\Theta = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$ is crucial because the EM algorithm converges towards a local maximum of the likelihood function. Usually it is recommended to use a random or uniform initial estimate for π and $\mathbf{A}$ [128]. An initial estimate of $\mathbf{B}$ is helpful for a DDHMM and necessary for a CDHMM. A possible estimate of $\mathbf{B}$ is to manual partition the observation sequences in states and then average over the observations for each state.

- **Choice of the Model:** Performance of an HMM also relies on the choice of the model. This considers the structure of the underlying Markov chain, number of states and model for the output.

The most popular application of HMMs is in acoustics and speech recognition [128, 129]. Further applications are handwriting recognition, protein sequence analysis, and human movement analysis [26, 86]. Lately, it has found application in gait identification.

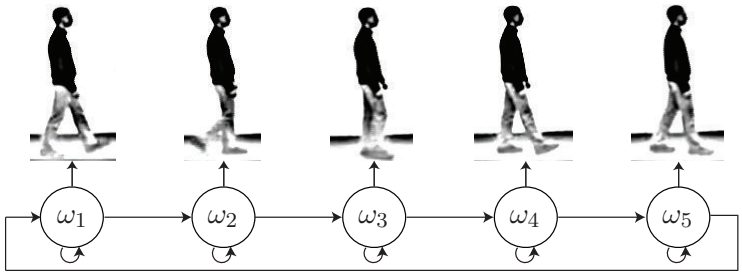


Fig. 6.2: The HMM transits the states ω_1 to ω_5 during a single step.

In vision-based gait analysis, a CDHMM is applied in [71] and good results are achieved for identification. This approach is adopted for marker-based gait analysis and compared with results of static classification, which is developed in chapter 5. It should be noted that further HMM concepts exist in vision-based gait analysis, such as factorial or parallel HMMs [23]. As the main purpose within this chapter is to investigate if modeling different stances is beneficial in marker-based gait analysis for affect recognition, the general concept of an HMM as described in [128] has been implemented and adopted. Further enhancements of the HMM such as parallel or factorial Markov chains may increase the recognition rate achieved with a CDHMM in marker-based gait analysis.

6.3 Results

This section summarizes the results for applying a CDHMM to recognize affect in gait patterns. The underlying principle is that a gait cycle is divided in several stances $\mathcal{E} = \{\mathbf{e}_1, ...\mathbf{e}_K\}$. A cyclic left-to-right HMM is used to model the transition from one stance to the next, see Fig. 6.2. Thus, the hidden state $\omega_i(t)$ represents the current stance and the continuous observations $\mathbf{V} = \{\mathbf{v}(1), ...\mathbf{v}(T)\}$ with $\mathbf{v} \in \mathbb{R}^d$ are the measured joint angles gained from marker-based recordings. To enable comparison with static classification, the same set of joint angles is used as in the previous chapter 5. In this case, the dimension of the observation vector $\mathbf{v}$ is $d = 22$. Training a CDHMM with an observation vector of this size requires a large number of training sequences and long computation time. Furthermore, computation of the loglikelihoods may result in values which exceed the precision range of the computer. This motivates to reduce the dimension of the observation vector. One possibility is to select a smaller number of joint angles. Furthermore, Kale *et al.* propose an algorithm which calculates the distance between the observation and each stance [71]. This distance vector is then used as observation vector of the CDHMM. In this case, the size of $\mathbf{v}$ is efficiently reduced to the number of stances K. For vision-based gait analysis, their method achieves good results, see [71, 162], and is used for marker-based gait analysis in the following.

In the following, results are reported for three different methods. In the first approach, the gait cycle is divided into a number of stances and a simple minimum distance classifier is applied. In a second approach, this method is extended by a transition matrix. This results in an CDHMM. The parameters of this CDHMM are iteratively trained and only a small number of joint angles are selected for the observation vector $\mathbf{v}$. In a third approach,

the method of Kale *et al.* [71] is adopted for marker-based affect recognition. In doing so, a distance metric is calculated which reduces the number of observations and a selection of a subset of joint angles is not required. Finally, these three approaches are compared, whereby the second and third approach differ in their representation of the observation vector.

A single stride of the gait sequences is extracted from the recordings for all three approaches. Thus after preprocessing, the input data is the same as in chapter 5.

6.3.1 Minimum Distance Classifier

The gait can be subdivided in a set of several stances $\mathcal{E} = \{\mathbf{e}_1, ..., \mathbf{e}_K\}$ with $\mathbf{e}_k \in \mathbb{R}^d$. In [71], a number of 5 stances is chosen for a single step. The feature vector in [71] contains the absolute width of the silhouette. Thus, it does not distinguish whether the right or the left leg is the support leg. In this case, the joint angles are different for the support and swing leg. This suggests to use a number of $K = 10$ stances for a complete stride. For comparison, the minimum distance classifier is also tested with half the number of stances. For the CDHMM later on, this will reduce the number of states and thus the complexity of the model. However, the stances are less precise.

The single stances $\mathbf{e}_k$ are estimated by dividing the training sequences into K equal parts and averaging over the joint angles within each part. Depending on the classification task, $\mathcal{E}^c$ is either a representative for a single walker c, for one affect c, or for one affect of one walker c. Thus, for identification 13 representatives $\mathcal{E}^c$ are trained, for inter-individual affect recognition 4, and for person-dependent emotion recognition 52.

In the simplest case, a minimum distance classifier can be chosen for decision making. The sample sequence $\boldsymbol{S} = \{\mathbf{s}(1), \mathbf{s}(2), ..., \mathbf{s}(T)\} \in \mathbb{R}^{T \times d}$ is compared with each of the representatives $\mathcal{E}^c$. The class for which the following criterion is minimized

$$\text{argmin}_c \sum_{t=1}^{T} \min_{k \in 1,...,K} D(\mathbf{s}(t), \mathbf{e}_k^c)$$

is assigned to the sample. The distance metric $D(\cdot, \cdot)$ is the Euclidean metric. The results for this approach are summarized for identification in Tab. 6.1 and for affect recognition in Tab. 6.2. Even though it shows excellent performance on identification for the neutral gait trials, recognition rates are lower if the walker expresses emotions during the identification process. Considering that the database includes all emotions, only 69% accuracy is achieved if the walking styles differ in expressed affect. The gait trials of four out of the 13 walkers are always assigned to the same but wrong walker. The accuracy is noticeable lower than the recognition rates in the previous chapter, see Tab. 5.5. Thus, the approach is susceptible to variations in the data. This is in accordance with [71] who report that the minimum distance classifier is susceptible to noise in gait analysis. Also, recognition rates for inter-individual and person-dependent recognition do not reach best rates of Tab. 5.3 and Tab. 5.4. Furthermore, similar recognition rates for a minimum distance classifier are achieved if only a single stance is used being equivalent to not dividing the gait in several stances. Thus, only dividing the gait in single stances is not sufficient to achieve proper

Stances	Identification [%] Data Set				
	Neut.	Happy	Sad	Angry	All
10	100	92	85	92	69
5	100	92	85	92	69
1	100	92	77	77	69

Tab. 6.1: Identification.

Stances	Recognition of Affect [%] Inter-ind.	Person-dep.
10	55	85
5	54	85
1	51	80

Tab. 6.2: Emotion Recognition.

classification of affect or identity.

Taking a closer look at the data shows that the recordings of the joint angles differ. Some trajectories of repeatedly recorded joint angles are more similar whereas others highly differ in range, offset and shape. From this follows that the minimum distance classifier seldom distinguishes between different stances and gives only a rough estimate which representative $\mathcal{E}^c$ the sample $\boldsymbol{S}$ is most similar to. The following two modifications are made to improve the performance of the minimum distance classifier:

- normalization of the joint angles,
- weighting of the joint angles.

The parameters of the standard normalization μ_a and σ_a are estimated from the training data for each joint angle a. The standard normalization is applied to the test data. Then, the range of the different joint angles is similar in average, still some joint angles vary more than others in shape and offset depending on the recording. This variation is not necessarily in correlation with the identity or affective state of the walker.

Fig. 6.3 illustrates how this affects the estimation of the stances $\mathbf{e}_k$. This motivates to introduce a weighting. The weights are separately calculated for the training set of each class. With $\boldsymbol{\sigma}_c(t)$ being the standard deviation of the samples $\boldsymbol{s}(t)$ belonging to class c, the mean of the standard deviations $\bar{\boldsymbol{s}}_c = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\sigma}_c(t)$ describes the average deviation between the trajectories. This measure is scaled by the absolute range of each angle a leading to

$$\tilde{w}_{a,c} = \left(\frac{\bar{s}_{a,c}}{s_{max,a,c} - s_{min,a,c}} \right)^{-2} ,$$

where the index a denotes the calculation for each angle. Large values of $\tilde{w}_{a,c}$ are desired for similar trajectories. Finally, the weights are normalized so that $w_{a,c} = \frac{\tilde{w}_{a,c}}{\sum_a w_{a,c}}$. Tab. 6.3 lists the angles with the five highest weights for each classification task. The listed weights are average values over the leave-one-out iterations. In all three cases, the angles of the spine, pelvis and foot progress are important whereas angles of the head, neck, elbow, and shoulder play a minor role. Unlike the selection of a subset of joint angles in chapter 5, here the joint angles of the lower body have higher weights. Thus, the following recognition rates are calculated by information provided to a larger extend by observation of the lower part of the body.

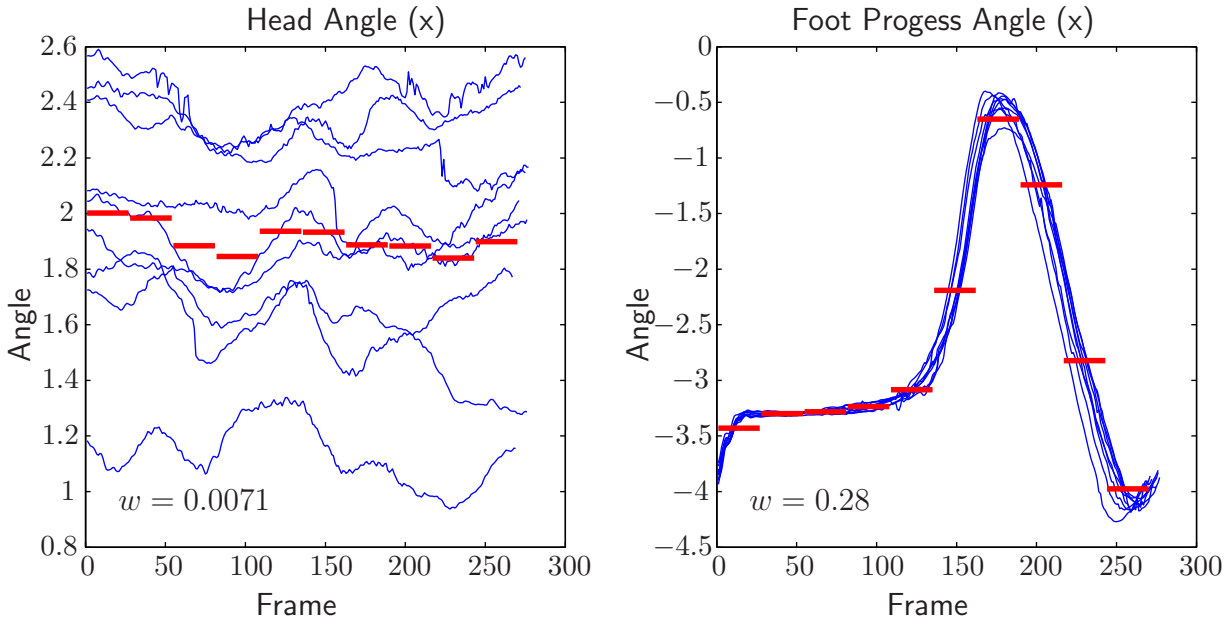


Fig. 6.3: The trajectories of the foot progress angle around the x-axis are very similar among the recordings, whereas the trajectories of the head angle around the x-axis differ strongly. High weights are assigned to angles for which the stances $\mathbf{e}_k$ are estimated usefully. The 10 red bars depict the stance value of each angle during the 10 intervals.

Identification		Affect: Inter-individual		Affect: Person-dependent	
Weight	Angle	Weight	Angle	Weight	Angle
.097	Foot Progress (X)	.25	Foot Progress (X)	.1	Foot Progress (X)
.07	Head (X)	.08	Spine (Y)	.1	Spine (Z)
.07	Spine (Z)	.08	Pelvis (Y)	.08	Foot Progress (Y)
.07	Foot Progress (Y)	.07	Shoulder (X)	.06	Pelvis (Y)
.07	Pelvis (Y)	.06	Spine (Z)	.05	Pelvis (Z)

Tab. 6.3: Weights for identification, inter-individual and person-dependent affect recognition.

The distance calculation of the minimum distance classifier is modified in the way that the difference $D(\cdot,\cdot)$ between the stance $\mathbf{e}_k^c$ and a sample $\boldsymbol{s}(t)$ is scaled for each angle a by $w_{a,c}$. Recognition rates are summarized in the Tables 6.4 - 6.7. Furthermore, the weights can be averaged over the classes c resulting in one common weight vector $\boldsymbol{w}$.

As it can be seen in the Tables 6.4 - 6.7, recognition rates are in general better when the weight vectors are averaged as when class dependent weights are used. Two points are worth to mention for the class dependent weights. First, it performs slightly better than the minimum distance classifier for identification but worse for affect recognition. Second, when the gait is not divided in several stances, recognition rates are only slightly above chance. The results for average weighting are interpreted in more detail further on. For identification, it can be seen in Tab. 6.6 that the minimum distance classifier is more robust against variations in the walking style caused by different affective states. A possible reason is that joint angles of the lower body part have higher weights and, hence,

Stances	Identification [%] Data Set Neut.	Happy	Sad	Angry	All
10	100	100	92	92	84
5	100	77	100	77	92
1	15	15	31	23	54

Tab. 6.4: Class-dependent weighting.

Stances	Recognition of Affect [%] Inter-ind.	Person-dep.
10	42	88
5	41	79
1	33	32

Tab. 6.5: Class-dependent weighting.

Stances	Identification [%] Data Set Neut.	Happy	Sad	Angry	All
10	100	92	92	100	92
5	100	92	92	100	100
1	100	85	100	77	92

Tab. 6.6: Average weighting.

Stances	Recognition of Affect [%] Inter-ind.	Person-dep.
10	56	93
5	55	91
1	48	72

Tab. 6.7: Average weighting.

changes in the body posture do less infer the classification. Even though, average weighting also concentrates on extracting information from the lower part of the body for affect recognition, it improves accuracy for inter-individual and person-dependent recognition. Results of person-dependent recognition are in the range of static classification with best feature extraction. Results for inter-individual affect recognition do not reach recognition rates achieved with static classification. Furthermore, dividing the gait in several stances is beneficial for the minimum distance classifier. Recognition rates vary only slightly when the number of stances K is 5 or 10. If only a single stance $K = 1$ is used, recognition rates decrease.

Summing up, the minimum distance classifier has been extended by a weighting which assigns high weights to gait trajectories which differ less among recordings. In doing so, the minimum distance classifier achieves similar recognition rates than static classification for identification and person-dependent affect recognition. Recognition rates for inter-individual affect recognition is lower than for static classification.

6.3.2 CDHMM based on a Set of Joint Angles

In the previous subsection, a minimum distance classifier estimates the most likely set of stances $\mathcal{E}^c$, which produces the observation sequence $\boldsymbol{V}$. The stances $\boldsymbol{e}_k^c$ of $\mathcal{E}^c$ are estimated by dividing the training set in K equal parts and averaging over the observed joint angles within each part. The transition matrix of a CDHMM determines a sequential order of the stances. In contrast to the minimum distance classifier, skipping stances is not possible and the duration of the stances can differ. An iterative training of the HMM estimates the transition matrix $\boldsymbol{A}^c$, the stances $\boldsymbol{e}_k^c$ and the observation distribution $b_j^c(\boldsymbol{v}) = \mathcal{N}(\boldsymbol{v}, \boldsymbol{e}_j^c, \boldsymbol{\Sigma}_j^c)$ for each state j. A CDHMM model with the parameter set Θ^c is trained for each class c. A number of 13 models is trained for identification, 4 models for inter-individual affect

recognition, and 52 models for person-dependent affect recognition. During the test phase, the probability $\log[P(\boldsymbol{V}|\Theta^c)]$ is calculated for all classes c that the model Θ^c generates the observed data $\boldsymbol{V}$. The model with highest probability is chosen and its class label is assigned to the test sample.

If the system is in state ω_j, the stance $\boldsymbol{e}_j$ is most probably observed. Thus the number of states J equals the number of stances K. During walking, the HMM transits from one state to the next in a sequential order. Thus, the transition matrix for $K = J = 5$ is

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & 0 & 0 & 0 \\ 0 & a_{22} & a_{23} & 0 & 0 \\ 0 & 0 & a_{33} & a_{34} & 0 \\ 0 & 0 & 0 & a_{44} & a_{45} \\ a_{51} & 0 & 0 & 0 & a_{55} \end{pmatrix} .$$

It is more probable that the system stays in the state ω_i than that the system transits to the next state ω_j; therefore, the property $a_{ii} > a_{ij}$ is valid $\forall i \neq j$. The observation $\boldsymbol{v}(t)$ is modeled by a multivariate normal density distribution

$$b_j^c(\boldsymbol{v}(t)) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}_j^c|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{v}(t) - \boldsymbol{e}_j^c)^T(\boldsymbol{\Sigma}_j^c)^{-1}(\boldsymbol{v}(t) - \boldsymbol{e}_j^c)\right] ,$$

where $\boldsymbol{\Sigma}_j^c$ is the covariance matrix of the stance j. A number of five joint angles Γ_2 is selected for $\boldsymbol{v}(t)$. The angles with highest weights in Tab. 6.3 are chosen for each classification task. In doing so, the same joint angles contribute to the classification task, which are most relevant for the minimum distance classifier with the weighted distance metric. This facilitates comparison of both approaches. For comparison with results in Chapter. 5.1.1, recognition rates are also calculated for the set of joint angles Γ_1 containing the neck forward angle, the shoulder flexion angle, the shoulder abduction angle, and the thorax forward angle. Tab. 6.8 lists the joint angles for each subset Γ_1 and Γ_2.

As a single stride beginning with the left leg approaching forward is extracted from the recorded sequences, the initial state distribution is $\boldsymbol{\pi} = [1, 0, ..., 0]^T$. The parameters $\boldsymbol{A}^c$, $\boldsymbol{e}_k^c$, and b_j^c are trained according the following re-estimation formulas which are implemented after the description of CDHMMs in [128]. As noted earlier, the parameters are trained for each class separately. For more comprehensible notation, the superscript c is neglected in the description of the training procedure. The re-estimation formulas require:

- the forward variable $\alpha_i(t) = P(\boldsymbol{v}(1)\boldsymbol{v}(2)...\boldsymbol{v}(t), \omega(t) = \omega_i|\Theta)$, which is the probability that the system is in state ω_i at time t and the partial observation sequence until t has been observed. The values $\alpha_i(t)$ are iteratively calculated by

$$\alpha_j(t+1) = \left[\sum_{i=1}^{K} \alpha_i(t)a_{ij}\right] b_j(\boldsymbol{v}(t)) \quad \text{for} \quad 1 \leq t \leq T-1, 1 \leq i \leq K ,$$

  with the initialization $\alpha_i(1) = \pi_i b_i(\boldsymbol{v}(1))$ for $1 \leq i \leq K$.

- the backward variable $\beta_i(t) = P(\boldsymbol{v}(t+1)\boldsymbol{v}(t+2)...\boldsymbol{v}(T), \omega(t) = \omega_i|\Theta)$, which is the probability of the partial observation sequence from time t to T and that the system

is in state ω_i at time t. The values $\beta_i(t)$ are iteratively calculated by

$$\beta_i(t) = \sum_{j=1}^{K} a_{ij} b_j(\boldsymbol{v}(t+1))\beta_j(t+1) \quad \text{for} \quad t = T-1, T-2, ..., 1, 1 \leq i \leq K \ ,$$

with the initialization $\beta_i(T) = 1$ for $1 \leq i \leq K$.

- a scaling, which prevents the variables to reach the precision range of the computer too fast. Rabiner [128] introduces the scaling factor c_t

$$c_t = \frac{1}{\sum_{i=1}^{K} \alpha_i(t)} \quad \text{and} \quad \hat{\alpha}_i(t) = c_T(t)\alpha_i(t), \quad \hat{\beta}_i(t) = c_t\beta_i(t) \ .$$

The re-estimation formulas are then applied to $\hat{\alpha}$ and $\hat{\beta}$. The loglikelihood that a observation $\boldsymbol{V}$ is generated by the model Θ is

$$\log[P(\boldsymbol{V}|\Theta)] = -\sum_{t=1}^{T} \log(c_t) \ .$$

- the variable $\gamma_i(t) = P(\omega(t) = \omega_i|\boldsymbol{V}, \Theta)$ which is the probability that the system is in state i at time t. It is calculated by

$$\gamma_i(t) = \frac{\hat{\alpha}_i(t)\hat{\beta}_i(t)}{\sum_{j=1}^{K} \hat{\alpha}_j(t)\hat{\beta}_j(t)} \ .$$

- initialization of $\boldsymbol{e}_i$ and $\boldsymbol{\Sigma}_i$. Each of the N training sequences is divided into K equal parts and the observations for each stance i are combined. Then, the mean stance $\boldsymbol{e}_i$ and the covariance matrix $\boldsymbol{\Sigma}_i$ are estimated for each stance i.

The initialization of $\boldsymbol{e}_i$ and $\boldsymbol{\Sigma}_i$ is conducted for each class c separately. In doing so, the iterative re-estimation starts near a local maxima of the loglikelihood $\log[P(\boldsymbol{V}|\Theta^c)]$. The re-estimation procedure allows that the $\log[P(\boldsymbol{V}|\Theta)]$ increases shortly to reach regions with a higher local maxima.

The parameters $\boldsymbol{A}^c$, $\boldsymbol{e}_k^c$, and b_j^c are iteratively re-estimated with a maximum number of five iterations. For a number of N observations, the re-estimation formulas are:

$$\begin{aligned}
\bar{a}_{ij} &= \frac{\sum_{n=1}^{N}\sum_{t=1}^{T} \hat{\alpha}_i^n(t) a_{ij} b_j(\boldsymbol{v}^n(t+1))\hat{\beta}_j^n(t+1)}{\sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{K} \hat{\alpha}_i^n(t) a_{ij} b_j(\boldsymbol{v}^n(t+1))\hat{\beta}_j^n(t+1)} \ , \\
\bar{\boldsymbol{e}}_i &= \frac{\sum_{n=1}^{N}\sum_{t=1}^{T} \gamma_i^n(t)\boldsymbol{v}^n(t)}{\sum_{n=1}^{N}\sum_{t=1}^{T} \gamma_i^n(t)} \ , \\
\bar{\boldsymbol{\Sigma}}_i &= \frac{\sum_{n=1}^{N}\sum_{t=1}^{T} \gamma_i^n(t)(\boldsymbol{v}^n(t) - \boldsymbol{e}_i)(\boldsymbol{v}^n(t) - \boldsymbol{e}_i)^T}{\sum_{n=1}^{N}\sum_{t=1}^{T} \gamma_i^n(t)} \ .
\end{aligned}$$

These equations assume that all observations $\boldsymbol{V}^n$ are equally probable and independent of each other. For dealing with dependent observations, the reader is referred to [95].

Angle Set	Recognition	Joint Angle	X	Y	Z
Γ_1	All	Neck	x		
		Shoulder	x	x	
		Thorax	x		
Γ_2	Identification	Head	x		
		Spine			x
		Pelvis		x	
		Foot Progress	x	x	
	Affect: inter-individual	Shoulder	x		
		Spine		x	x
		Pelvis		x	
		Foot Progress	x		
	Affect: person-dependent	Spine			x
		Pelvis		x	x
		Foot Progress	x	x	

Tab. 6.8: Selection of two subsets of joint angles for the CDHMM - Γ_1 is selected for comparison with static classification and Γ_2 for comparison with the minimum distance classifier.

	Identification [%]				
Stan-	**Data Set**				
ces	Neut.	Happy	Sad	Angry	All
10	54	54	62	54	69
5	85	54	54	62	69

Tab. 6.9: CDHMM with Γ_1.

	Recognition of Affect [%]	
Stances	**Inter-ind.**	**Person-dep.**
10	59	80
5	63	82

Tab. 6.10: CDHMM with Γ_1.

	Identification [%]				
Stan-	**Data Set**				
ces	Neut.	Happy	Sad	Angry	All
10	77	77	77	69	92
5	85	85	100	85	85

Tab. 6.11: CDHMM with Γ_2.

	Recognition of Affect [%]	
Stances	**Inter-ind.**	**Person-dep.**
10	40	72
5	38	72

Tab. 6.12: CDHMM with Γ_2.

After the parameters $\boldsymbol{A}^c$, $\boldsymbol{e}_k^c$, and $\boldsymbol{\Sigma}^c$ are estimated based on the training set, the CDHMMs are applied to the test set and the $\log[P(\boldsymbol{V}|\Theta^c)]$ is calculated for each test sequence. The results for the leave-one-out procedure are summarized in the Tables 6.9 - 6.12.

In all cases, recognition rates are above chance level but are less than best recognition rates achieved with static classification. For the subset of joint angles Γ_1, static classification with different dimension reduction techniques and the CDHMM rely on the same base data and therefore Γ_1 is suited for a more detailed comparison. For the static clas-

sification, minimum, mean, and maximum of each gait trajectory are calculated and for the CDHMM, the trajectories are modeled by a Markov Chain. Comparing the recognition rates of both approaches, shows that inter-individual recognition achieves the same rate of 63% for both approaches. Yet, recognition rates are lower for the CDHMM in the case of identification and person-dependent affect recognition. Afterward, the subset of joint angles Γ_1 and Γ_2 are compared with regard to recognition rates. The subset Γ_1 is better suited for recognition of affect and the subset Γ_2 better for identification. This is explained by the fact that Γ_1 contains joint angles which are mentioned to be significantly influenced by different affective states in related literature from psychology. On the other hand, Γ_2 contains mostly joint angles of the lower body part. Thus, it is easier to obtain information about the identity than information about the affective state from observing the lower body part. Thus for automatic recognition of emotions, it is useful to observe especially the upper part of the body during walking. Considering that also humans observe the upper body part of a walker more intensely during an emotion recognition task, see chapter 3.3.2, this procedure seems to be not only advantageous in machine learning but also for human recognition processes.

Finally, the performance of the CDHMM is compared with the minimum distance classifier. Including a Markov chain which models the transition from one stance to the next does not automatically improve recognition rates. As the number of joint angles need to be reduced for training the CDHMM, performance depends on the subset Γ. For inter-individual recognition of emotion, higher recognition rates are achieved with the CDHMM. But for person-dependent affect recognition, recognition rates are lower for the CDHMM. Identification with the CDHMM is for Γ_1 and Γ_2 better if the training set contains all trials and not only a subset of only neutral, happy, sad, or angry walking styles. In contrast to the minimum distance classifier, the CDHMM improves if more gait trials and different walking styles are available during training. Notwithstanding, recognition rates are lower for the CDHMM than for the minimum distance classifier with average weighting.

6.3.3 CDHMM based on the FED-Vector

Using the joint angles as the observation vector of the CDHMM has the disadvantage that only a small number of joint angles can be modeled. To overcome this limitation, Kale *et al.* propose a frame to exemplar distance (FED) [71]. Hereby, the observation vector $\boldsymbol{v}_{FED}(t)$ of the CDHMM is the distance $D(\cdot,\cdot)$ between the current observation $\boldsymbol{v}(t)$ and the stances $\boldsymbol{e}_i \in \{\boldsymbol{e}_1, ...\boldsymbol{e}_K\}$. Both $\boldsymbol{v}(t)$ and $\boldsymbol{e}_i$ contain a number of d joint angles. Kale *et al.* mention three distance measures, namely

- the Euclidean distance $D(\boldsymbol{v}(t), \boldsymbol{e}_i) = \|\boldsymbol{e}_i - \boldsymbol{v}(t)\|$,
- the inner product distance $D(\boldsymbol{v}(t), \boldsymbol{e}_i) = 1 - \frac{\boldsymbol{v}^T(t)\ \boldsymbol{e}_i}{\sqrt{\boldsymbol{v}^T(t)\boldsymbol{v}(t)\boldsymbol{e}_i^T\boldsymbol{e}_i}}$,
- and the sum of absolute distances $D(\boldsymbol{v}(t), \boldsymbol{e}_i) = \sum_{a=1}^{d} |v_a(t) - e_{a,i}|$, for which the index a denotes one of the d joint angles.

Fig. 6.4 illustrates calculations of the three distance metrics applied to one gait sequence. The distance $D(\boldsymbol{v}(t), \boldsymbol{e}_i)$ is minimal for the first stance $\boldsymbol{e}_1$ for all three metrics shortly

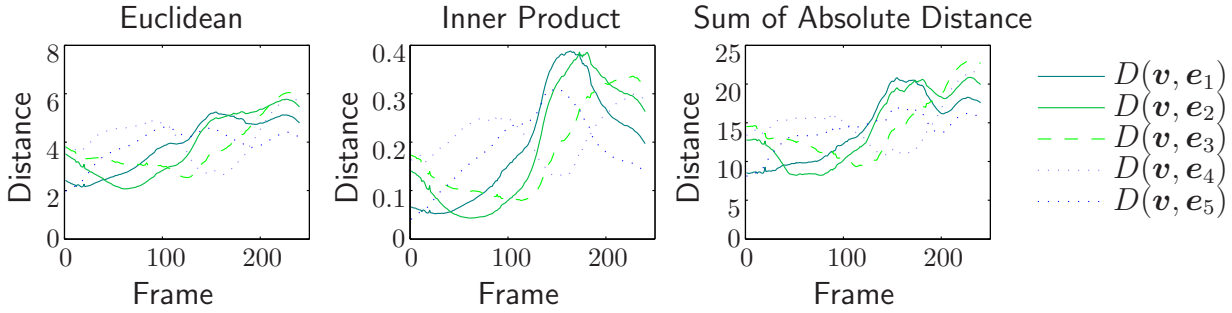


Fig. 6.4: During a stride, the CDHMM transits the states ω_1 till ω_K and in the same way the according distance metric $D(\boldsymbol{v}(t), \boldsymbol{e}_i)$ should be the smallest. $(K = 5)$

after the beginning of each stride. Stance $\boldsymbol{e}_K$ is similar to stance $\boldsymbol{e}_1$ so that it is possible that $D(\boldsymbol{v}(t), \boldsymbol{e}_K) < D(\boldsymbol{v}(t), \boldsymbol{e}_1)$ at the beginning, see e.g. the inner product distance in Fig. 6.4. With proceeding time, the distance becomes minimal for the subsequent stance $\boldsymbol{e}_i$. The FED vector $\boldsymbol{v}_{FED}(t) = [D(\boldsymbol{v}(t), \boldsymbol{e}_1), ..., D(\boldsymbol{v}(t), \boldsymbol{e}_K)]^T$ contains the distance metrics D for all K stances. The change of the maximum value of its entries over time reflects the transition from one stance to the next. In applying $\boldsymbol{v}_{FED}(t)$, the dimension of the observation vector is efficiently reduced from the number of angles d to the number of stances K.

If $\boldsymbol{v}_{FED}(t)$ is the observation of the CDHMM, the stances $\boldsymbol{e}_i$ are only used for the calculation of $\boldsymbol{v}_{FED}(t)$. The stances $\boldsymbol{e}_i$ are not reestimated during the training of the CDHMM for the indirect approach of applying the FED vector [71]. The observation distribution $(\boldsymbol{\mu}^C, \boldsymbol{\Sigma}^C)$ of the CDHMM models $\boldsymbol{v}_{FED}(t)$. The parameters $\boldsymbol{A}^C$, $\boldsymbol{\mu}^C$, and $\boldsymbol{\Sigma}^C$ of the CDHMM are trained in the same manner as described in the previous subsection 6.3.2. The achieved recognition rates using the Euclidean metric are reported in Tab. 6.8. The identity of the walker cannot be recognized using this algorithm and marker-based recordings. Person-dependent emotion recognition is better than chance level but performs worse than the other algorithms of this chapter.

Even though good performance of the FED vector is reported in vision-based gait analysis, a selection of a subset of joint angles performs better for marker-based gait analysis. A possible explanation therefor is that vision-based recordings contain information about the size and the contour of the walkers. This study concentrates on analyzing only joint angles, which are only implicitly affected by the size of the walker. Thus, the stance $\mathcal{E}$ of a walker contains information about the size and the contour in vision-based gait analysis, but in this approach the stance $\mathcal{E}$ contains only information about the kinematics. During a stride, the changes in the distance of the kinematics and each single stance $\boldsymbol{e}_i$ are larger than the difference of the kinematics and each single stance $\boldsymbol{e}_i$ between different walkers. Furthermore, considering the distance between markers of the elbow, shoulder, or knee as input vector, similar to the width vector in computer vision, leads to good performance of the FED approach, see [187]. This is a further hint that the FED approach requires information about the contour of the walker, which is missing if joint angles are used as input.

Stances	Identification [%] Data Set Neut.	Happy	Sad	Angry	All
10	0	0	0	8	8
5	8	15	31	15	15

Tab. 6.13: CDHMM with FED.

Stances	Recognition of Affect [%] Inter-ind.	Person-dep.
10	33	64
5	38	75

Tab. 6.14: CDHMM with FED.

6.4 Summary

In the previous chapter, statistics such as minimum, mean, and maximum of each stride are computed as features. In this chapter, the gait is subdivided in a number of K stances and a minimum distance classifier or an HMM are applied for classification. In the following, the results for these approaches are summarized and then compared to the recognition rates achieved with static classifiers. Afterwards general conclusions are drawn.

Classification with a minimum distance classifier requires normalization of the joint angles and a weighting. The weighting assigns higher weights to joint angles which vary less between gait trials. In this case, a minimum distance classifier achieves similar recognition rates for identification as a CDHMM with a set of selected joint angles as observations. For inter-individual and person-dependent recognition of affect, this CDHMM performs better than the minimum distance classifier with an increase of 10% in the recognition rate. Performance depends on the selected joint angles. A subset of joint angles, which is mentioned in related literature [32] to be significantly influenced by the affective state of the walker, achieves better recognition of affect than the subset of joint angles, which differ less amongst the recordings. Applying the distance vector circumvents a selection of joint angles. Yet, recognition rates are lower. The number of stances does affect the recognition rate, however it depends on the application whether five or ten stances are better to model a stride.

In comparison to static classifiers with efficient feature extraction, dividing the gait in several stances and applying a minimum distance classifier or an HMM does not result in higher recognition rates. Either achieved recognition rates for affect are in a similar range or lower. An explanation therefor would be that an HMM is robust against changes in a walker's speed. Yet, psychological studies have shown that even though speed is not the only factor, it is a relevant feature for emotion expression [32, 131]. Thus, classification with an HMM is comparable to scaling each gait sequence to a common reference speed. This would suggest that an HMM would perform better than a static classifier if the walking speed is changed during expressing the same affect.

Even though applying an HMM to gait analysis seems to be a promising approach, Boulgouris *et al.* mention in their comparison of algorithms for vision-based gait analysis that there might be a more efficient way to model gait dynamics [10]. This is in accordance with afore described results. It is concluded for marker-based gait analysis that the algorithms presented in chapter 5 are equal or better than the studied HMMs to recognize affect and additionally require less computational power.

6.5 Limitations

Dividing the gait in several stances seems to be advantageous for classification. Yet, the investigated techniques are not sufficient to improve recognition rate in comparison to static classification with efficient feature extraction. Developing more advanced algorithms which consider the dynamics of gait are a future research direction. A possibility would be to investigate parallel or factorial HMMs as in [23, 86].

Furthermore, this chapter focuses on the comparison of HMMs and static classifiers. The results of this comparison refer to the investigated techniques and the conclusions are limited to the these techniques. A general statement whether static or dynamic classification is better for gait analysis is not possible, as a larger number of techniques exist for both approaches. For example, static classification which models joint angle trajectories by eigenpostures in chapter 5.2.2, performs worse than an HMM. Studying only these two approaches would lead to the opposed conclusion. Thus, the result of the comparison relies on the investigated techniques. More advanced HMMs may perform better than the investigated techniques from static classification in chapter 5. Still, this chapter shows that the algorithms with efficient feature extraction proposed in chapter 5 perform well even if they do not divide the gait in several stances.

7 Expressions of Emotions in Gait Patterns for Robots

Humans interact socially with computers and robots [130]. To establish a natural human-robot interaction (HRI), it is not only required that the robot recognizes and understands social behavior, but also that it reacts in an appropriate manner. Humans recognize emotions from facial expressions, speech, gesture, and body motion during human-human interaction. Hence, these modalities can be used to design appropriate emotive reactions of the robot in HRI. Within this context, only a small number of robots are capable to express facial expressions or possess emotive speech processing. However, all robots are capable of link motions and some even of locomotion. This motivates to investigate whether emotions can be expressed in motions, such as walking. Therefore, this chapter focuses on the aspect whether a robot can express emotions in variations of the walking style. Hence, this chapter finishes the investigation of gait as modality for affective computing by exploring emotion expression whereas the previous chapter 3 analyzes human perception and the chapters 5 and 6 automatic emotion recognition. Possible application scenarios are personal assistance, service robotics, and entertainment.

Bethel and Murhy [8] introduce proximity zones to HRI. At this, body motions for emotion expression are especially suited for the personal and social proximity zone. Recent studies have shown that gesticulatory behavior of the robot and changes in its posture influences its appearance and support recognition of emotional states [102, 160]. This chapter focuses on locomotion as basic movement and investigates whether emotions are expressible in the way the robot walks. Experiments are conducted with a hexapod and an animated version of the hexapod. Human emotive gait patterns are recorded and are mapped to the robot. The gait patterns of the robot are finally evaluated. In particular, the following research questions are investigated:

1. How do human gait parameters, in particular step length, height, and time, differ regarding different emotional states?

2. Based on mapping these differences in the walking style to a robot, does the robot express emotions in gait such that they are sufficiently recognizable? Are different levels of pleasure, arousal, and dominance equally recognizable?

3. Are emotive gait patterns of a real robot or an animated robot perceived differently?

Results are that humans recognize emotions in the way the robot walks, differences in arousal and dominance are better expressed than differences in pleasure, and expressiveness between the robot and its animation differ only slightly. It should be noted that affective motions for robots in related studies are designed by humans. This approach focuses on a methodology which does not involve human adjustments and thus can be further adapted

for online, affective motion imitation and learning. The essential scientific contributions of this chapter have been previously published in [172].

The following section 7.1 gives an overview about designing emotional expressions for robots, particularly with regard to whole-body motions. First, section 7.2 analyzes statistically affective human gait patterns focusing barely on leg movement and then derives a mapping of the relevant features for implementation on a robotic platform. This mapping is evaluated in an experimental setup and the results are presented in section 7.3. The chapter ends with a summary in section 7.4 and limitations are discussed in section 7.5.

7.1 Emotion Expression in Body Motions

Emotions can be either studied in categories, e.g. Ekman's basic emotions [46], or as a point in a space, e.g. spanned by the axes pleasure, arousal, and dominance (PAD) [135]. The former is more related to basic linguistic usage, whereas the latter allows graduation of the emotional state in the emotion space and is more suited for technical applications. Both variants are applied for expression of emotions in literature.

Based on the facial action coding system (FACS), facial expressions have been implemented on robotic heads [13, 83]. However, this requires a sophisticated hardware. To bypass this, Hashimoto *et al.* propose a curved surface display to build a robotic head which is capable to express emotions [60]. For reliable emotion expression, body motion and facial expression should imply the same emotional state, otherwise the observer is left to guess [3]. Zecca *et al.* also show that humans recognize expressed emotional states of a humanoid better, if facial expressions are supported by the appropriate body postures [160]. Recognition rate is lower if the robot uses barely facial expressions and even lower if the robot uses barely emotive postures. The body postures have been designed by students, a cartoonist, and a photographer in this study.

Psychological studies indicate that emotions are not only expressed in gesticulatory behavior, but also in natural motions such as knocking, drinking and walking [32, 126, 131]. Several approaches have yet been undertaken to study emotional expressiveness in robotic motions [6, 100, 102, 113, 146]. Matsumara proposes a method based on Laban features as mathematical description for expression of emotions in postures [102]. The Laban features are derived from dance studies and quantify the two items effort and shape [55]. This approach is compared with discrimination using principal component analysis. Evaluation shows that the categories anger and joy have often been confused with each other as well as sad and fear. The motivation of Nakagawa *et al.'s* study is to find a general approach which modifies an arbitrary motion so that affective nuances are expressed [113]. They propose to divide an arbitrary motion in velocity and extensity of the motion and a basic posture. Velocity and extensity correlate with arousal and the basic posture relates to the expressed level of pleasure with a contracted posture for low pleasure and an open posture for high pleasure. Evaluation for the movements pointing and waving shows that the type of motion has an effect on the expressed emotion, and that the observers recognized the intended affective nuances except for the combination of high pleasure and low arousal. Also, Masuda *et al.* conclude from their study that the expressed emotion depends on the primary motion [100, 101]. They also utilize the Laban features to model the motions of

the robot. Takahashi *et al.* also designed the movements of the robot based on the Laban movement theory [146]. Additionally, they investigated the impact of the appearance of the robot on its expressiveness. Their results are that emotions can be already expressed with only a low degree of freedom robot, that the emotions are correctly expressed except for anger, and that the emotions anger and disgust depend on the appearance of the robot whereas the emotions joy, sadness, surprise, and fear do not depend on the appearance. The posture of the head is in detail studied in [6]. Beck *et al.* conclude that the head position is an important body posture and that changing a robot's head position is useful for non-verbal communication during HRI.

As most robots intended for human assistance will be equipped with a locomotion system, the following sections investigate whether a robot can also express emotions only in the way it walks.

7.2 Model

Mapping emotive motions from human to robot faces two basic challenges. First, the body structure of the robot differs from human physique both in kinematics and appearance. How does a mapping take this into account? Second, gait is a dynamic motion in comparison to posture. Consequently, the mapping has to consider modeling motion trajectories.

For designing social robots, Bar-Cohen and Breazeal suggest in [3] to adopt the nine principles of expressive animation which are described in [75, 149]. They describe how people interpret the motion of animated objects. In this context, conveying one emotional state at a time and exaggeration are of particular interest. Conveying one emotional state at a time increases the probability of a correct interpretation of the robot's behavior by a human observer. Thereby, it is beneficial if the robot expresses the same emotional state in all modalities to avoid misinterpretations and to increase authenticity, see chapter 3.3.3 for a related study with an animated puppet. Exaggeration of emotional expressions improves recognition [75, 149]. Currently, no design rules exist for exaggeration. Linear spatial exaggeration is applied for animated facial expressions in [54]. For mapping emotional gait patterns from human gait data to robot, this indicates that it is probably useful to exaggerate the relevant characteristics.

7.2.1 Analysis of Gait Parameters

Related literature shows that a person's emotional state affects her/his walking style , e.g. velocity is higher for angry and joyous gait than for neutral gait [32]. Within this study, the Munich gait database is analyzed for suitable gait characteristics which can be mapped to a walking robot. The data contains ten neutral gait patterns of each walker and ten gait trials during acting high or low pleasure, arousal or dominance. A set of relevant gait characteristics suited for mapping emotional gait patterns to a hexapod is investigated. These characteristics are step height h, step width w and cadence t.

Hypothesis 1: *If a suitable mapping can be found, gait patterns of a robot differing in step length, step height, and time express emotional states which are recognizable by a*

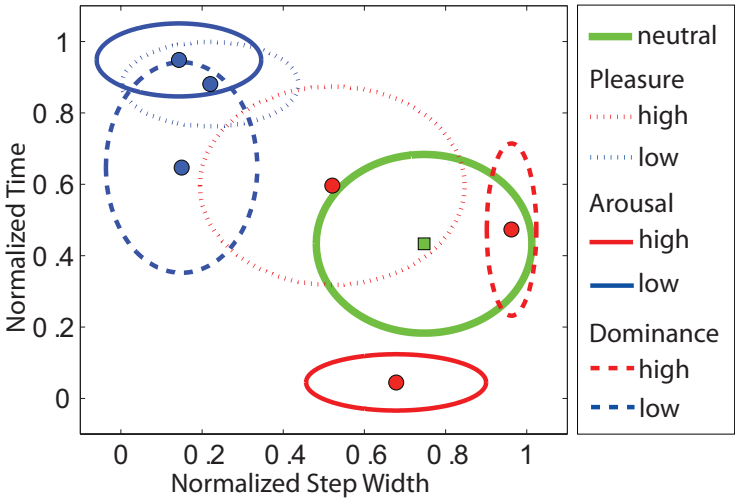


Fig. 7.1: Different walking styles form separable clusters in the normalized step width and normalized step time space. The clusters are approximated by an ellipse with mean values as centers and standard deviations as radii.

human observer. Within this study, the emotional states are separately analyzed for the dimensions pleasure, arousal, and dominance.

Fig.7.1 shows the distribution of the parameters step width $\overline{w}_{norm,emo}$ and cadence $\overline{t}_{norm,emo}$ for each emotion of the recorded walkers. The parameters vary between individuals; therefore they are normalized for each walker separately. The distribution of each emotion is estimated based on the normalized data of all walkers. The mean value and an ellipse defined by the standard deviations are plotted for each emotional state. The distributions of high and low arousal are well separable from each other. On the contrary, low and high pleasure are less separable and the clusters lie closer to each other. Furthermore, the distributions overlap, so that some emotional states are expressed similar and misinterpretations can occur between them. The neutral gait and the expression of high pleasure vary most among the walkers, as illustrated by high standard deviation for the normalized step width and time. Low as well as high arousal and dominance form smaller clusters. Thus, it is expected that these states are also better expressible in robotic gait patterns.

Hypotheses 2: *Differences in arousal and dominance are easier to retrieve from changes in robotic gait patterns than differences in pleasure based on mapping solely the parameters step width, height, and time.*

The following Figs. 7.2-7.3 illustrate the changes of each gait parameter comparing to the neutral gait pattern without normalization. Expression of high arousal, pleasure, or dominance often increases the parameters simultaneously, however the amplitude of increase differs. In a similar manner, low arousal, pleasure, or dominance decreases the parameters often simultaneously. The two exceptions are, 1) the noticeable increase of step length for high dominance, and 2) the increase of velocity for high arousal which is accompanied by a decrease in time and almost constant step length.

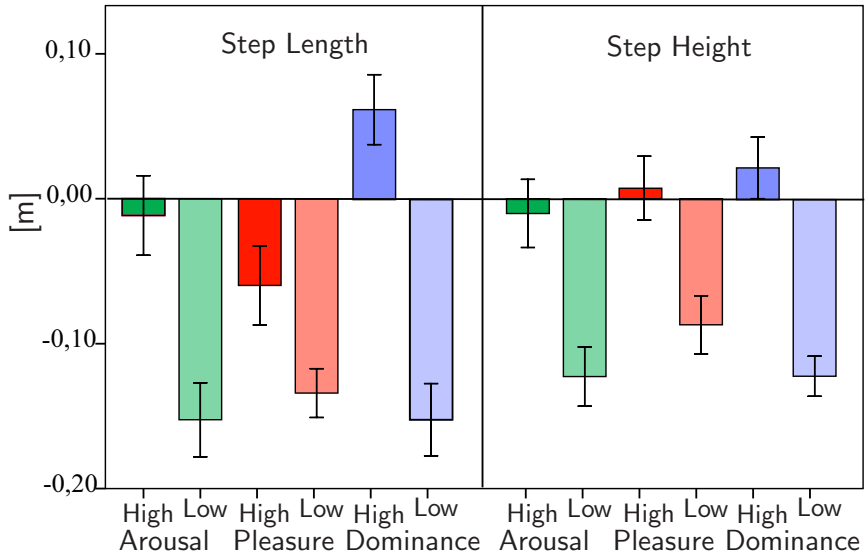


Fig. 7.2: Mean values and standard deviation of the parameters step length and height among all walkers differ for high or low expression of arousal, pleasure or dominance comparing to neutral gait.

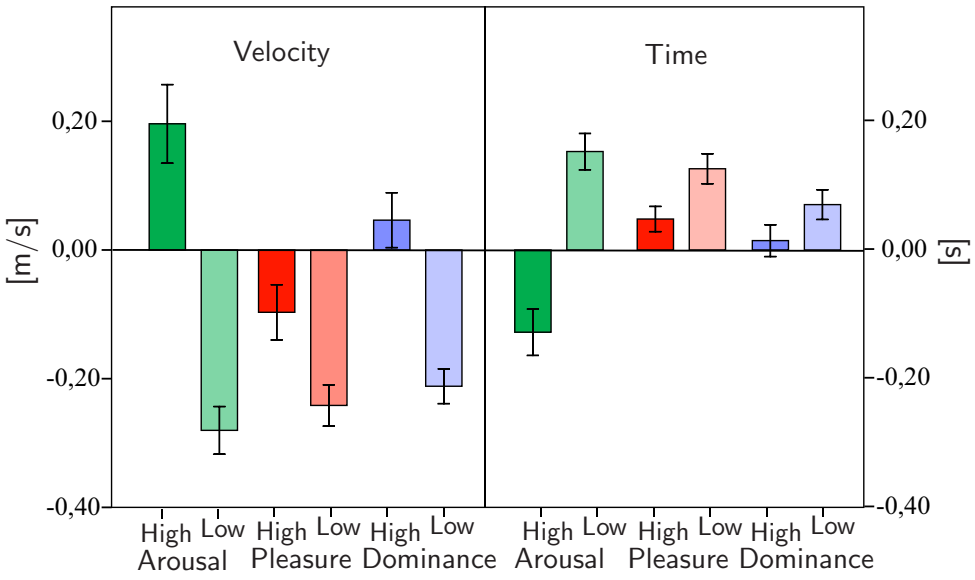


Fig. 7.3: Mean values and standard deviation of the parameters velocity and time among all walkers differ for high or low expression of arousal, pleasure or dominance comparing to neutral gait.

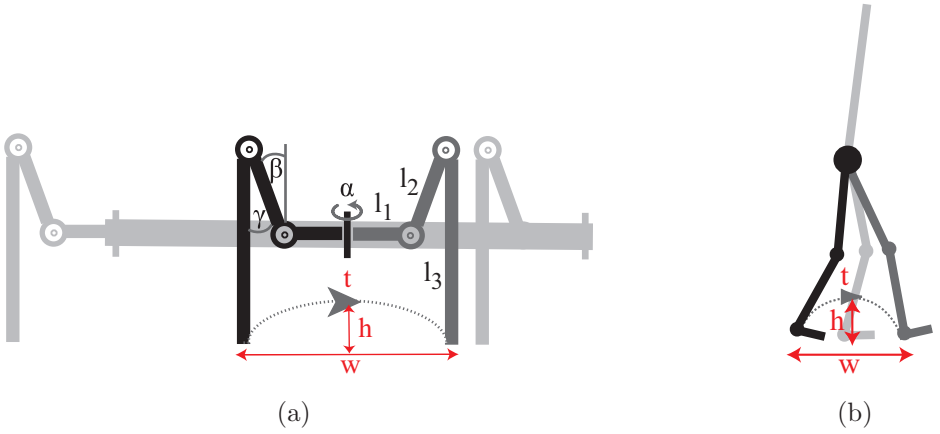


Fig. 7.4: Movement of one leg during a step is shown exemplarily for (a) the hexapod (black and dark gray: active leg) and (b) the human model (black and dark gray: swing leg).

7.2.2 Mapping of the Gait Parameters

Mapping of expressive motions from human to robot addresses 1) extraction of relevant parameters, 2) handling of different body structures, and 3) optional exaggeration. Within this study, a minimalistic approach is considered. Only the parameters which define the motion of the swing leg are mapped. Its trajectory is approximated by a parabola. The form of the parabola is defined by step width w and step height h. By mapping the time t for a single step, the gait velocity is set implicitly. Fig. 7.4(a) and 7.4(b) illustrate the motion of the swing leg modeled by the parameters step width w, height h and time t for a single step.

As walking robot, a hexapod has been chosen to avoid instability during gait cycles. A tripod gait variable in step length, height and time has been developed for the 18 DOF hexapod [184]. Each leg contains three servo-motors. The joints α, β, and γ are connected by the links l_1, l_2, and l_3. A microcontroller ATMega168 serves as servocontroller generating PWM signals to drive the servo-motors [1]. A second microcontroller ATMega324 calculates the sampled trajectories based on the gait parameters and handles communication with the PC and the servocontroller, for details see [184].

Except a limited number of highly sophisticated humanoids, physique between humans and robots differ. Therefore, gait parameters recorded from human emotive gait trials can not directly be applied to robots and a mapping is required. Within this work, the considered gait parameters affect only the lower body.

In biology and biomechanics, the Froude number describes dynamic similarity of walking styles, e.g. walking and running, among animals whose skeletons differ [103]. Equal Froude number indicates dynamic similarity. The Froude number F is defined as $F = \frac{v^2}{gl_h}$ with the gravitational acceleration g, the velocity v and the hip height l_h. If the robot walks at equal Froude number comparing to a human, it does so at a speed which is proportional to the square root of its hip height to human hip height. This approach was implemented on the hexapod. However, an experiment showed that the gait patterns of the robot are too similar between each other and differences were hardly recognizable [184]. Thus, this

Parameter	α	β	γ	Time
Minimum	5°	5°	5°	0.4 s
Maximum	30°	45°	45°	1.4 s

Tab. 7.1: Joint angle limitations of the hexapod.

method inspired from biology is not suitable for this task and it is concluded that emotive robotic gait patterns need to be expressed more distinguishably and demand exaggeration.

In a second approach, normalized human gait parameters are mapped to the hexapod by defining lower and upper limits of each parameter given by hardware limitations. The minimum and maximum values of each joint are listed in Tab. 7.2.2. This method automatically includes exaggeration, because human emotive gait parameters vary in a small range comparing to maximum possible range. In the following, the procedure is generally described for the parameter $y \in \{w, h, t\}$ and $n = 13$ walkers:

1. Calculate minimum $y_{i,min}$ and maximum $y_{i,max}$ of all trials of walker $i = 1...n$.
2. Calculate the normalized parameter $y_{norm,emo}$ for each emotion and each walker i

$$y_{norm,i,emo} = \frac{y_{i,emo} - y_{i,min}}{y_{i,max} - y_{i,min}} \ . \tag{7.1}$$

3. Average $y_{norm,i,emo}$ over all walkers $\overline{y}_{norm,emo} = \frac{1}{n} \sum_{i=1}^{n} y_{norm,i,emo}$.
4. Calculate the corresponding parameter of the hexapod $y_{hexa,emo}$ based on $\overline{y}_{norm,emo}$

$$y_{hexa,emo} = \overline{y}_{norm,emo}(y_{hexa,max} - y_{hexa,min}) + y_{hexa,min} \ . \tag{7.2}$$

5. Derive joint angle from the parameter $y_{hexa,emo}$.

This procedure leads to step height $h_{hexa,emo}$, step length $w_{hexa,emo}$ and the time for one step $t_{hexa,emo}$ of the hexapod. Applying inverse kinematics provides the desired joint angles:

$$\beta_{emo} = \arcsin\left([\overline{h}_{norm,emo}(\sin\beta_{max} - \sin\beta_{min}) + \sin\beta_{min}\right] \tag{7.3}$$

$$\alpha_{emo} = \arcsin\left[\left(\frac{\overline{w}_{norm,emo}(w_{hexa,max} - w_{hexa,min}) + w_{hexa,min})}{2(l_1 + l_2\cos\beta_{emo})}\right]\right. \tag{7.4}$$

with

$$w_{hexa,max} = 2\sin\alpha_{max}(l_1 + l_2\cos\beta_{min}) \tag{7.5}$$

$$w_{hexa,min} = 2\sin\alpha_{min}(l_1 + l_2\cos\beta_{max}) \ . \tag{7.6}$$

The lower leg is perpendicular to the ground for the swing legs and the support legs so that

$$\gamma_{emo} = \beta_{emo} \ . \tag{7.7}$$

Normalized Parameter	Neutral	Pleasure		Arousal		Dominance	
		low	high	low	high	low	high
$\overline{w}_{norm,emo}$	0.75	0.52	0.22	0.68	0.14	0.96	0.15
$\overline{h}_{norm,emo}$	0.71	0.73	0.31	0.68	0.16	0.82	0.17
$\overline{t}_{norm,emo}$	0.43	0.60	0.88	0.05	0.95	0.47	0.65

Tab. 7.2: Normalized gait parameters.

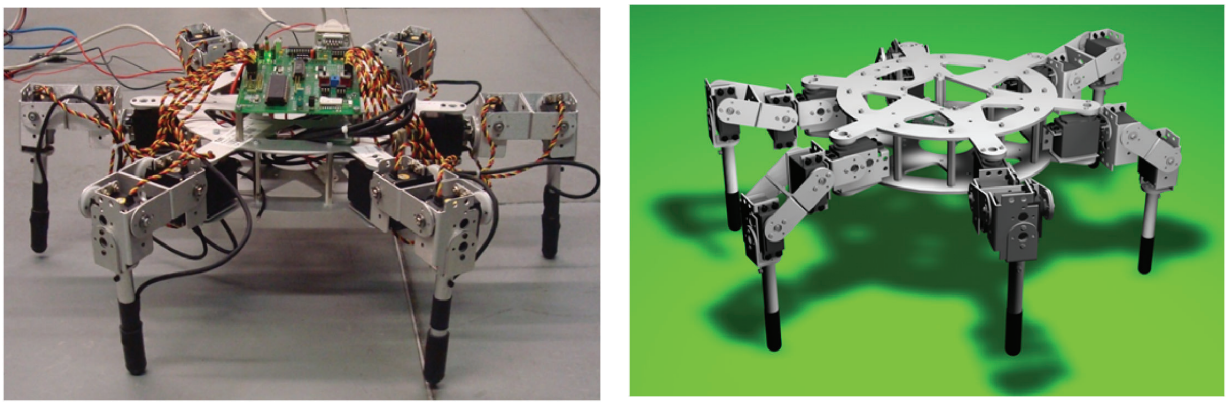


Fig. 7.5: Photo of the hexapod and snapshot of its animation.

The swing legs belonging to the active tripod move equidistantly from $-\alpha_{emo}$ to α_{emo} covering the distance $w_{hexa,emo}$. The vertical motion describes a parabolic arc starting at $-\beta_{emo}$, reaching the maximum $h_{hexa,emo}$ with β_{emo} and ending again in $-\beta_{emo}$. The joints of the passive tripod $\beta_{passive,emo} = -\beta_{emo}$ define the height of the platform.

The normalized parameters $\overline{w}_{norm,emo}$, $\overline{h}_{norm,emo}$, and $\overline{t}_{norm,emo}$ are listed in Tab. 7.2.2 for each emotion. The joint angles are derived from these parameters according to Eqns. 7.2-7.6.

It is known that real robots seem to be more encouraging and advising during interaction than their animated 3D-visualizations [127, 144]. Still, to the authors knowledge, no study has yet compared the impression of expressive motions for a real robot versus its animation. To analyze, whether also the expression and perception of emotions is affected by the representation of the robot - real versus animated -, an animation of the hexapod has been constructed [183]. The walking styles were implemented on the real hexapod and on its animation, see Fig. 7.5. In both cases, the robot walks in a 45°angle towards the observer.

Hypotheses 3:*Although the same kinematics are applied to the robot and its animation, differences in the perceived emotions are expected due to higher authenticity of the real robot.*

7.3 Experiment

The algorithms have been evaluated in a psychological experiment with 24 participants (19 male, 5 female, mean age: 25.8 ± 3.2) [181]. The gait patterns have been transferred to an

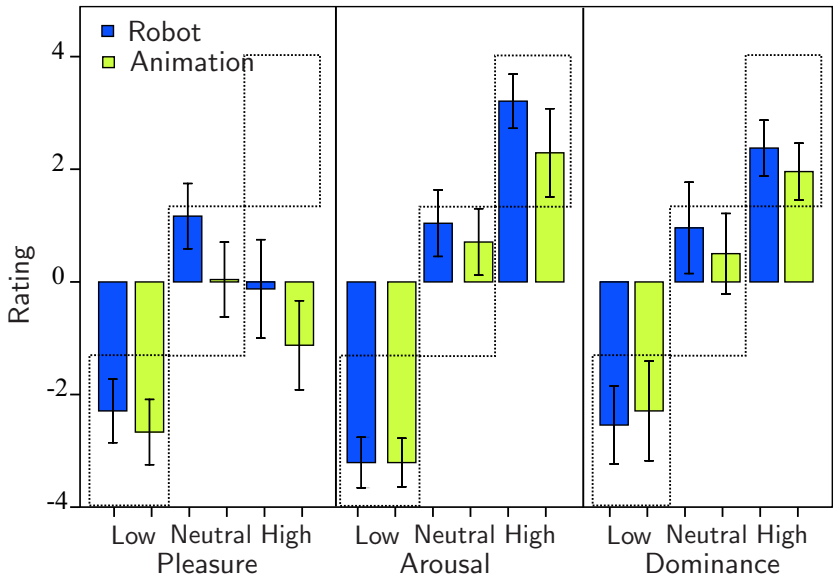


Fig. 7.6: Participants' ratings of the gait patterns lie in the expected areas for each level on each dimension except for high pleasure.

animation of a hexapod and a real robot, see Fig.7.5. Participants rated the expressed affect of the animation and the robot separately for each affective dimension on a 9-item Likert scale $(-4\ , ..., +4)$. Order of the level of expressed affective dimension and order of the affective dimension was randomized among the subjects. Starting with either animation or real robot was balanced among the subjects.

Fig. 7.6 summarizes the results for the evaluation of emotive gait patterns expressed by the hexapod. Mean values of the ratings of all participants are plotted and the error bars represent the 95% confidence interval. Expected regions for each level are marked by a dashed rectangle. The intended levels for low, medium, and high pleasure, arousal, or dominance were recognized on average by the participants except for high pleasure. This result holds for both the animation and the robot. However, within each region, the robot was generally rated to be more pleasant, more active, and more dominant than its animation. Furthermore, the results show that the neutral gait pattern is slightly too dominant and too active. This is traced back to the fact, that the neutral region of the recorded gait patterns lies closer to the clusters of high arousal and high dominance, see Fig. 7.1. A less expressible neutral gait is desirable for future work.

Tab. 7.3 lists the ratings of each participant. The ratings $(-4, -3, -2)$ are assigned to low expression, $(-1, 0, 1)$ to neutral expression and $(2, 3, 4)$ to high expression. Best recognition is achieved for different levels of arousal with a rate of 86% for the animation of the hexapod and 85% for the robot. Recognition rates for distinguishing three different levels of dominance are 69% for the animation and 68% for the robot, respectively. Recognition rates for pleasure are lower with 54% for the animation and 47% for the robot, respectively. Chance level of 33% is exceeded in all cases. Highest recognition is achieved for arousal followed by dominance. Gait patterns expressing high pleasure are often confused with neutral or even low pleasure; therefore, recognition for the dimension pleasure is low. Furthermore, recognition rates for arousal and dominance are equal for

Intended Level	Recognized Level								
	Pleasure			Arousal			Dominance		
Animation	L	N	H	L	N	H	L	N	H
L(ow)	**20**	4	0	**23**	1	0	**20**	1	3
N(eutral)	4	**16**	4	1	**19**	4	3	**15**	6
H(igh)	12	9	**3**	1	3	**20**	0	9	**15**
Robot	L	N	H	L	N	H	L	N	H
L(ow)	**19**	5	0	**23**	1	0	**20**	3	1
N(eutral)	1	**11**	12	0	**17**	7	4	**10**	10
H(igh)	6	14	**4**	0	3	**21**	0	5	**19**

Tab. 7.3: Participants' ratings summarized in a confusion matrix.

the robot and its animation. Thus, expressiveness of the robot and the animation does not differ if expression of each dimension is divided in only three labels 'low', 'neutral', or 'high'. As three values on the rating scale were combined to a single label, which are 'low', 'neutral', or 'high', tendencies within a region can not be detected by comparing the calculated recognition rates.

Subsequently, a repeated measures analysis of variance (ANOVA) is applied for each dimension separately [181]. The within-subject factors are level of expression and style, which is either robot or animation. Ratings for different levels of pleasure differ, $F(2, 46) = 41.87$, $p = .00$, $\eta_p^2 = .65$, and style has a minor effect, $F(1, 23) = 10.52$, $p = .04$, $\eta_p^2 = .31$. Following pairwise comparisons are significant at $p < .05$ for correct level of pleasure. However, high pleasure is perceived falsely as between neutral and low pleasure. Expression of pleasure for the animation is rated lower in comparison to the robot, $p < .004$.

A large effect is assessed for different levels of activity, $F(2, 46) = 246.10$, $p = .00$, $\eta_p^2 = .92$, and a minor effect for style, $F(1, 23) =$, $p = .03$, $\eta_p^2 = .20$. Following pairwise comparisons are significant at $p < .05$ for correct recognized levels of arousal in all cases. The animation appears to be less active than the robot, $p < .03$.

Also, the gait patterns are perceived differently depending on the degree of dominance, $F(1.69, 38.92) = 54.29$, $p = .00$, $\eta_p^2 = .70$ (Huynh-Feldt correction). No significant effect is observed for style, $F(1, 23) = 1.80$, $p = .19$, $\eta_p^2 = .07$. Following pairwise comparisons are significant at $p < .05$ for correct level of dominance in all cases.

In summary, interpretation of the mean values, investigation of the confusion matrix, and analysis of variance leads to the following conclusions. Regarding *hypothesis 1*, human observers recognize different emotional states in the walking style of the robot if the parameters step length, step height and time are varied. Regarding *hypothesis 2*, different levels of arousal and dominance are better expressed and thus better recognized than different levels of pleasure in robotic gait patterns based on mapping step length, step height and time. Regarding *hypothesis 3*, emotion expression of the robot and its animation do not differ, considering only the levels 'low', 'neutral' and 'high' expression for a dimension. However, considering also slight tendencies, the results indicate that the robot appears to be more active and pleasant than its animation.

Lastly, participants were asked which dimension was easier to rate on a 9-item Likert scale. Different levels of activity were easier to estimate than different levels of dominance, which themselves were easier than different levels of pleasure, $F(2, 46) = 23.30$, $p = .00$, $\eta_p^2 < .50$. Half of the participants found it easier to recognize emotions from gait patterns for the robot than for the animation. Five participants preferred the animation and seven reported no difference. Given reasons for the preference of the real robot are sounds of the servo-motors, noise during ground contact, and presence. On the other hand, noise has been perceived as disrupting for most who preferred the animation.

7.4 Summary

Integration of emotions enhances naturalness of human-robot interaction (HRI). This requires that the robot is equipped with hardware to express emotions. Believability and recognition of expressions is increased if the same emotional state is expressed in all modalities which the robot is capable of. Within this aspect, this chapter analyzes if a walking robot can express emotions in the way it walks and if these expressions are recognizable.

The emotive gait patterns are derived from human characteristics for emotive gait. The parameters step length, height, and time vary depending on the emotion. Mapping these changes to the kinematics of the robot and exaggeration of the walking styles leads to distinguishable expressions for the dimensions pleasure, arousal, and dominance. Especially differences in arousal and dominance are easier to express. This has been expected because the original data of human emotive gait patterns recorded with motion capture differ stronger for differences in arousal and dominance than for pleasure. It is also in accordance with Matsumaru's study considering body postures of a teddy bear robot [102]. In this, the emotions anger and happiness differing highly in pleasure are often confused with each other. This is observable in further studies, so that it is suggested that body motions, except gestures, are more suited to express the level of activation or intensity of an emotion [47, 126]. Although robots appear to be more engaging than their animation [127], ratings of the gait patterns regarding emotions differ only in their tendency between the robot and its animation. The robot tends to be more pleasant and active than its animation.

This study shows that by changing its walking style the hexapod expresses emotions, in particular differences in arousal, which can be used to increase expressiveness in HRI. Biologically inspired mapping based on the Froude number did not result in recognizable emotions. Therefore, the range of motion of the robot has been extended to maximal range given by hardware limitations. This is comparable to linear exaggeration of the motion patterns.

In conclusion, the walking style of a robot influences how humans perceive its appearance, especially for the affective dimensions activity and dominance. Considering the study about the interrelation of facial expressions and walking styles for an animated puppet in chapter 3.3.3, adapting the walking style of the robot can increase authenticity and the probability to recognize emotive expressions of a robot in HRI.

7.5 Limitations

Within this study, the neutral gait pattern is expressed slightly too pleasant, active, and dominant. This is traced back to a disadvantage of the applied mapping. The mapping is based on normalization which requires a minimum and maximum for each parameter. These values have been taken from hardware limitations of the robot. However, hardware limitations differ between robots and even between robots of similar structure. An estimation of a gait pattern which expresses the neutral emotional state is desired independent of the robot's structure. Within this aspect, this approach has been tested only on a single robot. For generalization, a further step is required which extends the current approach to different robot structures, e.g. a humanoid or wheeled robot.

Furthermore, the dimensions pleasure, arousal, and dominance have been evaluated separately. Still, the gait parameters vary similar for different levels on the dimensions. Increasing the separability for each dimension is a relevant aspect for future work.

As this work focuses on gait as modality for affective computing, only the motion gait has been investigated. Yet, further motions of robots can be investigated regarding their emotional expressiveness. Within this context, the interaction of motion and posture is also of interest. Expressiveness can be increased by changing the posture such as bending of the upper body for displeasure. In general, developing a general framework for emotion expression in motions is advantageous to designing emotive movements individually for each motion.

8 Conclusions and Future Directions

Nonverbal communication will play a major role in future social HRI. This includes that the robot understands a human's emotional behavior and responds in an appropriate manner during the interaction. Yet to date, affective behavior is predominantly studied in psychology for human-human interaction. The research field affective computing studies human-robot and human-computer interaction which contains emotional interaction. This comprises research about how a robot recognizes human emotions, understands and reasons about emotional behavior, and expresses itself emotions. It is expected that including nonverbal communication to HRI increases natural interaction and is more convenient and entertaining. The capability to socially interact is a valuable supplement for robots sold as assistive, daily helpers.

Within the research field of affective computing, facial expressions and speech are studied as modality in the majority of cases. These modalities may not always be available or observable. Considering that humans express emotions also via other modalities such as body movements and gestures, studying a broad range of modalities is of interest in affective computing. Additionally, multi-modal emotion expression plays an important role to communicate irony, jokes, and authenticity. Furthermore, the term emotion is subdivided in several categories which are either recognized or expressed. Even though a dimensional emotion model has advantages for technical applications, it has yet less been investigated in affective computing [56].

This work studies the suitability of the daily, human motion gait as modality for affective computing. Advantages of gait as modality are that it provides a means to detect emotions at distance, it is less susceptible to deliberate social editing, it does not require an interaction, and the observation is non-intrusive in comparison to physiological parameters. This study highlights various aspects regarding recognition and expression of emotions in gait patterns.

8.1 Concluding Remarks

The primary task of walking is locomotion. On account of this, emotion expression in gait is covered by a cyclic walking movement which is affected by various factors such as physique, weight, complaints, and age. To filter out the information about the affective state of the walker from his/her walking behavior is a challenging task both for humans and machine learning algorithms.

From reviewing psychological studies in chapter 3, it is concluded that humans recognize different emotional categories in gait patterns above chance. Here, this work additionally shows that differences in arousal and dominance are better recognized than differences in pleasure. Thus, applying the dimensional PAD emotion model for automatic gait analysis seems to be beneficial. If different walking styles are combined with facial expressions,

authenticity increases if both modalities express the same emotion. Inappropriate walking styles influence the perception and credibility of other modalities like facial expressions. Thus, it is worthwhile to investigate expressive walking styles for robots.

In biomechanics and medicine, inferential statistics is applied traditionally for gait analysis. Recently, different machine learning approaches have been introduced. Chapter 4 compares inferential with predictive statistics. Mathematical relations are derived, in particular a way to estimate classification rates from published, statistical values of medical, psychological, or biomechanical studies. This contribution is not only relevant for gait analysis but also for other topics in which the data can be analyzed with inferential or predictive statistics in e.g. bioinformatics or affective computing.

Chapter 5 develops feature extraction techniques for different static classifiers while chapter 6 utilizes an HMM to model the transition between gait stances. Even though, the latter approach divides the gait cycle in a number of stances, it does not outperform static classification with efficient feature extraction. Feature selection based on psychological studies, PCA, LDA, and their kernel extensions are compared for feature extraction. Naive Bayes, NN, and SVM are applied for classification. If the number of training samples is small and the feature vector is highly dimensional, it is derived that the decision borders of NN coincide with the decision borders of a SVM with a Gaussian kernel. Results are that 1) emotions are recognized comparable to human performance, 2) including information about the identity of the walker in training the classifier increases the recognition rate, 3) differences in arousal and dominance are better recognized than differences in pleasure, and 4) the identity or the gender are easier to recognize in gait than the emotion. The algorithms are developed for marker-based gait analysis and use the joint angles during a single stride as input. Results serve as upper limit what can be achieved with vision-based gait analysis for the modality gait if computer vision algorithms reconstruct joint angles from 2D images with high accuracy. Recording the human gait in natural setups with normal cameras would also facilitate the recording of spontaneous emotions. In conclusion, retrieving the affective state from walking is possible but pattern recognition algorithms demand high accuracy of the recorded joint angles.

Chapter 7 examines whether changes in the walking style of a robot are sufficient to convey different emotions. At this, a mapping is derived which maps human changes in gait parameters to a robot. The obtained walking styles are implemented on a real and virtual hexapod and are evaluated by human observers in an experiment. The results are that differences in arousal and dominance are easier to recognize than differences in pleasure. Furthermore, the real robot tends to appear more active and pleasant than its animation. Therefore, a robot can convey emotions via changing its walking style. Yet, this requires a hardware which allows a sufficient operating range for each joint angle, and emotion expression is not as differentiable as for facial expressions or speech.

In summary it can be concluded, that the modality gait conveys emotions which are detectable with techniques from machine learning above chance level and expressible in robotic gait patterns. Also, the PAD model is well suited to study emotions in gait patterns besides categories of emotions. Limitations are that the human gait is highly individual, that different emotions are less distinct than for other modalities, and that high accuracy of the recorded joint angles is required for emotion recognition.

8.2 Outlook

Emotions influence a human's cognition, decision making, and behavior. During an interaction, nonverbal communication conveys valuable information about importance, validity, and social meaning via emotions. Retrieving and using this information in HRI is a challenging research field and will play a major role in social HRI. With its interdisciplinary orientation, it combines computer science and engineering with psychology and raises a number of research questions.

Affective computing is divided in automatic recognition of emotions, expression of emotions for robots and virtual characters, and modeling affective HRI interaction. Up to date, the modalities facial expression and speech have been predominantly studied. A smaller number of studies investigates physiological parameters, gesture, and body movements. This work extends the state of the art by analyzing the usability of the daily motion gait for affective computing. The modalities provide communication channels for nonverbal interaction and, hence, studying them is essential for social HRI. A broad range of modalities facilitates nonverbal communication during various situations and for different robot configurations. Furthermore, a number of studies have integrated emotional behavior in HRI and HCI for several applications.Within this context, it is expected that including nonverbal communication in HRI increases both natural interaction and entertainment while it decreases misunderstandings at the same time.

Future research may be directed towards the following issues which are especially relevant for integration of affective computing in HRI.

- Regarding the modality gait, the next step is to analyze spontaneous emotion expression. Markerless recording and vision-based gait analysis is therefor advantageous. Furthermore, extension to a broader range of body motions is relevant and increases applicability of observing human motions for recognition of affect. This concerns both automatic emotion recognition and emotional expressiveness of robots.

- Systems based upon a single modality are feasible for well-defined scenarios. In more flexible and dynamic environments, multimodal systems enhance recognition performance and research on the combination of different modalities is of special relevance. Within this context, crucial issues are the choice of the communication channels, fusing of the data carried by multiple channels, temporal aspects of the data from different channels, representation of the term emotion, automatic segmentation, and sensitivity to the context.

- A number of studies have been carried out which integrate affective behavior in HRI. At this, future aspects are modeling the interaction with stochastic mathematical models, developing concepts for choosing appropriate reaction strategies for the robot, and constructing common benchmarks for evaluation of the quality of the social interaction. A further aspect refers to ethical issues such as protecting a person's privacy.

In conclusion, affective computing is a fascinating research area within social robotics and finds application in assistive robotics.

Bibliography

[1] http://www.lynxmotion.com/.

[2] S. Abe. *Support Vector Machines for Pattern Classification*. Springer, 2005.

[3] Y. Bar-Cohen and C. Breazeal. *Biologically inspired intelligent robots*. Spie Press, 2003.

[4] J. Bates. The role of emotion in believable agents. *Communications of the ACM*, 37(7):122–125, Jul. 1994.

[5] G. Baudat and F. Anouar. Generalized discriminant analysis using a kernel approach. *Neural Computation*, 12:2385–2404, 2000.

[6] A. Beck, L. Canamero, and K. Bard. Towards an affect space for robots to display emotional body language. In *Proc. IEEE Int. Symp. Robot and Human Interactive Communication*, pages 491–496, 2010.

[7] D. Bernhardt and P. Robinson. Detecting affect from non-stylised body motions. In *Proc. of Int. Conf. Affective Computing and Intelligent Interaction, LNCS 4738*, pages 59–70, 2007.

[8] C. Bethel and R. Murphy. Survey of non-facial/non-verbal affective expressions for appearance-constrained robots. *IEEE Transactions on Systems, Man and Cybernetics - Part C: Applications and Reviews*, 38(1):83–92, Jan. 2008.

[9] C.M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[10] N.V. Boulgouris, D. Hatzinakos, and K.N. Plataniotis. Gait recognition: A challenging signal processing technology for biometric identification. *IEEE Signal Processing Magazine*, 22(6):78–90, Nov. 2005.

[11] G. Box and G. Tiao. *Bayesian Inference in Statistical Analysis*. John Wiley & Sons, 1973.

[12] C. Breazeal, A. Brooks, J. Gray, M. Hancher, J. McBean, W.D. Stiehl, and J. Strickon. Interactive robot theatre. *Communications of the ACM*, 46(7):76–85, 2003.

[13] C.L. Breazeal. *Designing Sociable Robots*. MIT Press, 2002.

[14] S. Brownlow, A. Dixon, C. Egbert, and R. Radcliffe. Perception of movement and dancer characteristics from point light displays of dance. *Psychological Record*, 47:411–421, 1997.

[15] D. Buller, J. Burgoon, C. White, and A. Ebesu. Interpersonal deception: Vii. behavioural profiles of falsification, equivocation and concealment. *Journal of language an social pdychology*, 13:366 – 395, 1994.

[16] C. Busso and S.S. Narayanan. Interrelation between speech and facial gestures in emotional utterances: a single subject study. *IEEE Transactions on Audio, Speech and Language Processing*, 8(8):2331–2347, November 2007.

[17] A.J. Calder, A.D. Lawrence, and A.W. Young. Neuropsychology of fear and loathing. *Nature Neuroscience*, 2:352 – 363, May 2001.

[18] A.C. Calvo and S. D'Mello. Affect detection: an interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, 1(1):18–37, 2010.

[19] W.B. Cannon. Against the James-Lange and the thalamic theories of emotions. *Psycho. Rev.*, 34:281 – 295, 1931.

[20] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. 2001. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[21] T. Chau. A review of analytical techniques for gait data: Part 1: Fuzzy, statistical and fractal methods. *Gait and Posture*, 13:49–66, 2001.

[22] T. Chau. A review of analytical techniques for gait data: Part 2: Neural network and wavelet methods. *Gait and Posture*, 13:102–120, 2001.

[23] C. Chen, J. Liang, H. Zhao, H. Hu, and J. Tian. Factorial HMM and parallel HMM for gait recognition. *IEEE Transactions on Systems, Man and Cybernetics - Part C: Applications and Reviews*, 39(1):114–123, Jan. 2009.

[24] L. Chen, H. Liao, M. Ko, J. Lin, and G. Yu. A new lda-based face recognition system which can solve the small sample size problem. *Pattern Recognition*, 33:1713–1726, 2000.

[25] R.E. Cochran, F.J. Lee, and E. Chown. Modeling emotion: Arousal's impact on memory. In *28th Annual Conference of the Cognitive Science Society*, pages 1133 – 1138, 2006.

[26] A. Cohen. Hidden markov models in biomedical signal processing. In *Proc. of 20th Int. Conf. IEEE Engineering in Medicine and Biology Society*, pages 1145–1150, 1998.

[27] J. Cohen. *Statistical Power Analysis for Behavioral Sciences*. Hillsdale, New Yersey: Lawrence Erlbaum Associates, 1977.

[28] J. Cohen. Things i have learned so far. *American Psychologist*, 45(12):1304–1312, 1990.

[29] J. Cohen. The earth is round (p<.05). *American Psychologist*, 49(12):997–1003, 1994.

[30] J. M. Cortina and H. Nouri. *Effect Size for ANOVA Designs*. Sage Publications, 2000.

[31] M. Coulson. Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. *Nonverbal Behavior*, 28(2):117–139, 2004.

[32] E. Crane and M. Gross. Motion capture and emotion: Affect detection in whole body movement. In *Proc. of Int. Conf. Affective Computing and Intelligent Interaction, LNCS 4738*, pages 95–101, 2007.

[33] C. Curio, M. Giese, M. Breidt, M. Kleiner, and H. Bülthoff. Exploring human dynamic facial expression recognition with animation. In *Proc. of Int. Conf. Cognitive Systems*, pages 161–166, 2008.

[34] J. Cutting and L. Kozlowski. Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of Psychonomic Society*, 9:353–356, 1977.

[35] T. Dalgleish. The emotional brain. *Nature*, 5:582–589, july 2004.

[36] Antonio R. Damasio. *Descartes' Error: Emotion, Reason and the Human Brain*. New York: Avon, 1994.

[37] C. Darwin. *The expression of emotions in man and animals*. Chicago university press, Chicago, 1872 1965.

[38] G.J. de Vries, P. Lemmens, and D. Brokken. Same or different? recollection of empathizing with an emotional event from the perspective of appraisal models. In *Proc. of Int. Conf. Affective Computing and Intelligent Interaction*, 2009.

[39] W. Deng, J. Hu, and J. Guo. Gabor feature based classification using LDA/QZ algorithm for face recognition. In *Proc. of Computer Analysis of Images and Patterns, LNCS 4221*, pages 15–24, 2006.

[40] W. Dittrich, T. Troscianko, and S. Lea ans D. Morgan. Perception of emotion from dynamic point-light displays represented in dance. *Perception*, 25:727–738, 1996.

[41] R.J. Dolan. Emotion cognition and behavior. *Science*, 298:1191 – 1194, 2002.

[42] P. Domingos and M. Pazzani. Beyond independence: Conditions for the optimality of the simple Bayesian classifier. In *In Proc. of 13th Int. Conf. on Machine Learning*, pages 105–112, 1996.

[43] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. John Wiley & Sons, 2001.

[44] W. Dunlap, J. Cortina, J. Vaslow, and M. Burke. Meta-analysis of experiments with matched groups or repeated measured design. *Psychological Methods*, 1:170–177, 1996.

[45] P. Ekman. Darwin, perception, and facial expression. *Annals of the New York Academy of Sciences*, 1000:105–221, 2003.

[46] P. Ekman and W. V. Friesen. A new pan-cultural facial expression of emotion. *Motivation and Emotion*, 10:159–168, Jun. 1986.

[47] P. Ekman and W.V. Friesen. Detecting deception from the body or face. *Journal of Personality and Social Psychology*, 29:288–298, 1974.

[48] S. Finch, G. Cumming, and N. Thomason. Colloquium on effect sizes: the roles of editors, textbook authors, and the publication manual: Reporting of statistical inference in the journal of applied psychology: Little evidence of reform. *Educational and Psychological Measurement*, 61(2):181–210, 2001.

[49] R. Fisher. Statistical methods and scientific induction. *Journal of Royal Statistical Society, Series B*, 17:69–78, 1955.

[50] R. Fisher. *Statistical methods and scientific inference*. Edinburgh: Oliver and Boyd, 1956.

[51] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 1990.

[52] B. Gelder. Toward the neurobiology of emotional body language. *Nature*, 7:242 – 249, March 2006.

[53] B. Gelder and N. Hadjikhani. Non-conscious recognition of emotional body language. *Sensory and Motor Systems*, 17(6):583–586, Apr. 2006.

[54] M. Giese, B. Knappmeyer, and H. Bülthoff. *Biologically Motivated Computer Vision*, chapter Automatic Synthesis of Sequences of Human Movements by Linear Combination of Learned Example Patterns, pages 538–547. Springer, LNCS 2525, 2002.

[55] A. Guest. *Labanotation: The System of Analyzing and Recording Movement*. Routledge, 2004.

[56] H. Gunes and M. Pantic. Automatic, dimensional and continuous emotion recognition. *International Journal of Synthetic Emotions*, 1(1):69 – 99, 2010.

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

[58] B. Guo and M. Nixon. Gait feature subset selection by mutual information. *IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans*, 39(1):36–46, Jan. 2009.

[59] S. Halovic and C. Kroos. Facilitating the perception of anger and fear in male and female walkers. In *Proc. of AISB, Symposium on Mental States, Emotions and their Embodiment*, 2009.

[60] M. Hashimoto, H. Kondo, and Y. Tamatsu. Effect of emotional expression to gaze guidance using a face robot. In *Proc. IEEE Int. Symp. Robot and Human Interactive Communication*, pages 95–100, 2008.

[61] T. Hastie, R. Tibshirani, and J Friedman. *The elements of statistic learning: data mining , inference and prediction*. Springer, 2009.

[62] A. Heberlein, R. Adolphs, D. Trannel, and H. Damasio. Cortical regions for judgements of emotions and personality traits from point-light walkers. *Cognitive Neuroscience*, 16:1143–1158, 2004.

[63] J.E. Hoffman. *Attention*, chapter 3: Visual Attention and Eye Movement, pages 119–121. London: Psychology Press Ltd, 1998.

[64] R. Huang, Q. Liu, H. Lu, and S. Ma. Solving the small sample size problem of LDA. In *Proc. of 16th Int. Conf. Pattern Recognition*, volume 3, pages 29–32, 2002.

[65] C.J. Huberty and S. Olejnik. *Applied MANOVA and discriminant analysis*. Willey Series in Probability and Statistics, 2006.

[66] I. Inza, P. Larranaga, R. Blanco, and A. Cerrolaza. Filter versus wrapper gene selection approaches in DNA microarray domains. *Artificial Intelligence in Medicine*, 31:91–103, 2003.

[67] P. Jafari and F. Azuaje. An assessment of recently published gene expression data analyses: reporting experimental design and statistical factors. *BMC Medical Informatics and Decision Making*, 6(27), 2006.

[68] W. James. What is an emotion? *Mind*, 9:188 –205, 1884.

[69] D. Janssen, W.I. Schöllhorn, J. Lubienetzki, K. Fölling, H. Kokenge, and K. Davids. Recognition of emotions in gait patterns by means of artificial neural nets. *Nonverbal Behavior*, 32(2):79–92, June 2008.

[70] G. Johansson. Visual perception of biological motion and a model for its analysis. *Perception ans Psychophysics*, 14(2):201–211, 1973.

[71] A. Kale, A. Sundaresan, A.N. Rajagopalan, N.P. Cuntoor, A.K. Roy-Chowdhury, V. Krüeger, and R. Chellappa. Identification of humans using gait. *IEEE Transactions on Image Processing*, 13(9):1163–1173, Sept. 2004.

[72] M. Kantardzic. *Data Mining: Concepts, Models, Methods, and Algorithms*. Wiley-IEEE Press, 2002.

[73] A. Kapur, N. Virji-Babul, G. Tzanetakis, and P.F. Driessen. Gesture-based affective computing on motion capture data. In *Proc. of Int. Conf. Affective Computing and Intelligent Interaction, LNCS 3784*, pages 1–7, 2005.

[74] C.D. Katsis, N. Katertsidis, G. Ganiatsas, and D.I. Fotiadis. 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, May 2008.

[75] I. Kerlow. *The Art of 3D Computer Animation and Effects*. Wiley & Sons, 2004.

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

[77] R.E. Kirk. Practical significance: a concept whose time has come. *Educational and Psychological Measurement*, 56:746–759, 1996.

[78] R.E. Kirk. Promoting good statistical practices: Some suggestions. *Educational and Psychological Measurement*, 61(2):213–218, 2001.

[79] A. Kleinsmith and N. Bianchi-Berthouze. Recognizing affective dimensions from body posture. In *In Proc. of Affective Computing and Intelligent Interaction, LNCS 4738*, 48-58, 2007.

[80] L.J. Kleinsmith and S. Kaplan. The interaction of arousal and recall interval in nonsense syllable paired associate learning. *Experimental Psychology*, 67:124 –126, 1964.

[81] G. Kleiter. *Bayes Statistik -Grundlagen und Anwendungen*. de Gruyter, 1980.

[82] L. Kozlowski and J. Cutting. Recognizing the sex of a walker from a dynamic point-light display. *Perception and Psychophysics*, 21:575–580, 1977.

[83] K. Kühnlenz, S. Sosnowski, and M. Buss. *Human-Robot Interaction*, chapter 12: Evaluating Emotion Expressing Robots in Affective Space, pages 235–245. I-Tech Education and Publishing, 2007.

[84] D. Kulic. *Safety for Human-Robot Interaction*. PhD thesis, The University of British Columbia, 2005.

[85] D. Kulic and E.A. Croft. Affective state estimation for human-robot-interaction. *IEEE Transactions on Robotics*, 23(5):991–1000, Oct. 2007.

[86] D. Kulic, W. Takano, and Y. Nakamura. Incremental learning, clustering and hierarchy formation of whole body motion patterns using adaptive hidden markov chains. *The International Journal of Robotics Research*, 27(7):761–784, July 2008.

[87] C. Lampert. Kernel methods for object recognition. Tutorial at the DAGM, 2008.

[88] P.J. Lang. Behavioral treatment and bio-behavioral assessment: Computer applications. *Technology in mental health care delivery systems*, pages 119–l37, 1980. Norwood, NJ: Ablex.

[89] J. Lasseter. Priniciples of traditional animation applied to 3D computer animation. In *Proc. of 14th Int. Conf. Computer Graphics and Interactive Techniques SIGGRAPH*, pages 35–44, 1987.

[90] R.S. Lazarus. *Emotion & Adaptation*. Oxford University Press, 1991.

[91] L. Lee and W.E.L. Grimson. Gait analysis for recognition and classification. In *Proc. of IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 148–155, 2002.

[92] R.V. Lenth. Some practical guidelines for effective sample-size determination. *The American Statistician*, 55(3):187–193, 2001.

[93] R.W. Levenson. *Social psychophysiology and emotion: Theory and clinical applications*, chapter Emotion and the autonomic nervous system: A prospectus for research on autonomic specificity, pages 17–42. New York: John Wiley & Sons, 1988.

[94] I. Levner. Feature selection and nearest centroid classification for protein mass spectrometry. *BMC Bioinformatics*, 6:68, 2005.

[95] X. Li, M. Parizeau, and R. Plamondon. Training hidden markov models with multiple observations - a combinatorial method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(4):371–377, Apr. 2000.

[96] Y. Li, J. Kittler, and J. Matas. Effective implementation of linear discriminant analysis for face recognition and verification. In *Proc. of 8th Int. Conf. Computer Analysis of Images and Patterns, LNCS 1689*, pages 234–242, 1999.

[97] H. Liu, J. Li, and L. Wong. A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns. *Genome Informatics*, 13:51–60, 2002.

[98] Y. Ma, H.M. Paterson, and F.E. Pollick. A motion capture library for the study of identity, gender and emotion perception from biological motion. *Behaviour Research Methods*, 38(1):134–141, 2006.

[99] A.M. Martinez and A.C. Kak. PCA versus LDA. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(2):228–233, Feb. 2001.

[100] M. Masuda and S. Kato. Motion rendering system for emotion expression of human form robots based on laban movement analysis. In *Proc. IEEE Int. Symp. Robot and Human Interactive Communication*, pages 344–349, 2010.

[101] M. Masuda, S. Kato, and H. Itoh. Emotion detection from body motion of human form robot based on laban movement analysis. In *Proc. of Int. Conf. Principles of Practice is Multi-Agent Systems, LNAI 5925*, pages 322–334, 2009.

[102] T. Matsumaru. Discrimination of emotion from movement and addition of emotion in movement to improve human-coexistence robot's personal affinity. In *Proc. IEEE Int. Symp. Robot and Human Interactive Communication*, pages 387–394, 2009.

[103] R. Mc Neill Alexander. *Principles of Animal Locomotion*. Princeton University Press, 2003.

[104] A. Mehrabian. Pleasure-arousal-dominance: a general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4):261 – 292, 1996.

[105] T.O. Meservy, M.L. Jensen, J. Kruse, J.K. Burgoon, j.F. Nunamaker, D.P.Twitchell, G. Tsechpenakis, and D.N. Metaxas. Deception detection trhough automatic, unobtrusive analysis of nonverbal behavior. *IEEE Intelligent Systems*, 20(5):36–43, Sep. 2005.

[106] J. Michalak, N. Troje, J. Fischer, P. Vollmar, T. Heidenreich, and D. Schulte. Embodiment of sadness and depression - gait patterns associated with dysphoric mood. *Psychosomatic Medicine*, 71:580–587, 2009.

[107] J.A. Mikels, B.L. Fredrickson, G.R. Larkin, C.M. Lindberg, S. Magold, and P.A. Reuter-Lorenz. Emotional category data on images from the international affective picture system. *Behavior Research*, 37(4):626–630, 2005.

[108] G.A. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63:81–97, 1956.

[109] T.B. Moeslund, A. Hilton, and V. Krüger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104(2-3):90–126, Nov.-Dec. 2006.

[110] J.M. Montepare, S.B. Goldstein, and A. Clausen. The identification of emotions from gait information. *Nonverbal Behavior*, 11:33–42, 1987.

[111] J.D. Morris. Observations: SAM: The Self-Assessment Manikin - an efficient cross-cultural measurement of emotional response. *Journal of Advertising Research*, 35, 1995.

[112] S.A. Mulaik, N.S. Raje, and R.A. Harshman. *What if there were no significance tests?*, chapter 4: There is a time and place for significance testing, pages 65–115. Psychology Press, 2009.

[113] K. Nakagawa, K. Shinozawa, H. Ishiguro, T. Akimoto, and N. Hagita. Motion modification method to control affective nuances for robots. In *Proc. of Int. Conf. Intelligent Robots and Systems*, pages 5003–5008, 2009.

[114] S. Nakagawa and I.C. Cuthill. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*, 82:591–605, 2007.

[115] J. Neyman. *First course in probability and statistics*. New York: Holt, Rinehart and Winston, 1950.

[116] P.M. Niedenthal. Embodying emotion. *Science*, 316:1002–1005, 2007.

[117] M.S. Nixon, T.N. Tan, and R. Chellappa. *Human Identification based on Gait*. Springer, 2006.

[118] M. Pantic and L. J. M. Rothkrantz. Automatic analysis of faial expressions: The sate of the art. *Transactions on Pattern Analysis and Machine Intelligence*, 22:1424–1445, December 2000.

[119] M. Pantic and L. J. M. Rothkrantz. Toward an affect-sensitive mutlimodal human-computer interaction. *Proceeding of the IEEE*, 9:1370–1390, September 2003.

[120] E.S. Pearson. Some thoughts of statistical inference. *Annals of Mathematical Statistics*, 33:394–403, 1962.

[121] A. Pecchinenda. The affective significance of skin conductance activity during a difficult problem-solving task. *Cognition & Emotion*, 10(5):481–504, 1996.

[122] C. Peter and R. Beale, editors. *Affect and Emotion in Human-Computer Interaction*, volume 4868 of *Lecture Notes in Computer Science*. Springer, 2008.

[123] R.W. Picard. *Affective Computing*. MIT Press, 1997.

[124] R.W. Picard. Affective computing: From laughter to ieee. *IEEE Transactions on Affective Computing*, 1(1):11–17, 2010.

[125] R.W. Picard, E. Vyzas, and J Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10:1175–1191, October 2001.

[126] F.E. Pollick, H.M. Paterson, A. Bruderlin, and A.J. Sanford. Perceiving affect from arm movement. *Cognition*, 82(2):B51–B61, Dec. 2001.

[127] A. Powers, S. Kiesler, S. Fussell, and C. Torrey. Comparing a computer agent with a humanoid robot. In *Proc. of ACM/IEEE Int. Conf. Human-Robot Interaction (HRI)*, 2007.

[128] L.R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, Feb. 1989.

[129] L.R. Rabiner and B.-H. Juang. *Fundamentals of Speech Recognition*. Prentice Hall, 1993.

[130] B. Reeves and C. Nass. *The media equation: how people treat computers, television, and new media like real people and places*. New York: Cambridge University Press, 1996.

[131] C.L. Roether, L. Omlor, A. Christensen, and M.A. Giese. Critical features for the perception of emotion from gait. *Journal of Vision*, 9(6):1–32, 2009.

[132] C.L. Roether, L. Omlor, and M.A. Giese. Lateral asymmetry of bodily emotion expression. *Current Biology*, 18(9):329–330, 2008.

[133] E. Rolls. *Emotion Explained*. Oxford University Press, 2007.

[134] J.A. Russell. A circumplex model of affect. *Personality and social psychology*, 39:1161 – 1178, 1980.

[135] J.A. Russell and A. Mehrabian. Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3):273–94, Sep. 1977.

[136] L. Sachs and J. Hedderich. *Angewandte Statistik: Methodensammlung mit R*. Springer, 2009.

[137] H. Sackeim, R. Gur, and M. Saucy. Emotions are expressed more intensely on the left side of the face. *Science*, 202(27):434–435, Oct. 1978.

[138] Y. Saeys, I. Inza, and P. Larranaga. A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23(19):2507–2517, 2007.

[139] S. Sarkar, P.J. Phillips, Z. Liu, I.R. Vega, P. Grother, and K.W. Bowyer. The humanID gait challenge problem: Data sets, performance, and analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(2):162–177, Feb. 2005.

[140] C. Scharfenberger, S. Chakraborty, and G. Färber. Robust image processing for an omnidirectional camera-based smart car door. *ACM Transactions on Embedded Computing Systems (to appear)*, 2011.

[141] K.S. Scherer. *Handbook of Cognition and Emotion*, chapter Appraisal Theory, pages 637–663. Wiley, New York, 1999.

[142] B. Schölkopf, A. Smola, and K.-R. Müller. Kernel principal component analysis. In *Proc. of Int. Conf. on Artificial Neural Networks*, pages 583–588, 1997, LNCS 1327.

[143] D.J. Sheskin. *Handbook of parametric and nonparametric statistical procedures*. Chapman & Hall/CRC, 4th edition, 2007.

[144] K. Shinozawa and J. Yamato. *Human-Robot Interaction*, chapter 19: Effect of Robot and Screen Agent Recommendations on Human Decision-Making, pages 346–356. I-Tech Education and Publishing, 2007.

[145] S.T. Skidmore and B. Thompson. Statistical techniques used in published articles: A historical review of reviews. *Educational and Psychological Measurement*, 70(5):777–795, 2010.

[146] K. Takahashi, M. Hosokawa, and M. Hashimoto. In *Proc. of IEEE Int. Conf. on Industrial Technology*.

[147] P. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining*. Pearson Addison Wesley, 2006.

[148] J.G. Taylor and N.F. Fragopanagos. The interaction of attention and emotion. *Neural Networks*, 18:353–369, 2005.

[149] F. Thomas and O. Johnston. *Disney Animation: The Illusion of Life*. Abbeville Press, New York, 1981.

[150] N. Troje. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision*, 2(5):371–387, 2002.

[151] L. van der Maaten, E. Postma, and H. van den Herik. Dimensionality reduction: a comparative review. Technical report, Tilburg University, 2009.

[152] VICON. Plug-in gait product guide - foundation notes.
H.G. Wallbott. Bodily expression of emotion. *European Journal of Social Psychology*, 28:880–896, 1998.

[153] H.G. Wallbott. Bodily expressions of emotion. *European Journal of Social Psychology*, 28:879–896, 1998.

[154] Y. Wang and L. Guan. Recognising human emotional state from audiovisual signals. *IEEE Transactions on Multimedia*, 5(5):936 – 946, August 2008.

[155] D.H. Wolpert and W.G. Macready. No free lunch theorem for optimization. *IEEE Transaction on Evolutionary Computation*, 1(1):67–82, Apr. 1997.

[156] J. Wu, J. Wang, and L. Liu. Feature extraction via KPCA for classification of gait patterns. *Human Movement Science*, 26(3):393–411, June 2007.

[157] T. Xu. *Aspects of Visual Attention for Autonomous Mobile Robots*. PhD thesis, TU München, 2010.

[158] R.M. Yerkes and J.D. Dodson. The relation of strength of stimulous to rapidity of habit formation. *Comparative neurology and psychology*, 18:459 –482, 1908.

[159] S. Yu, T. Tan, K.Huang, K. Jia, and X. Wu. A study on gait-based gender classification. *IEEE Transactions on Image Processing*, 18(8):1905–1910, Aug. 2009.

[160] M. Zecca, Y. Mizoguchi, K. Endo, F. Iida, Y. Kawabata, N. Endo, K. Itoh, and A. Takanishi. Whole body emotion expression for KOBIAN humanoid robot - preliminary experiments with different emotional patterns. In *Proc. IEEE Int. Symp. Robot and Human Interactive Communication*, pages 381–386, 2009.

[161] Z. Zeng, M. Pantic, G.I. Roisman, and T.S. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1):39–58, Jan. 2009.

[162] D. Zhang, Y. Wang, and B. Bhanu. Age classification based on gait using hmm. In *Proc. of Int. Conf. on Pattern Recognition*, pages 3834–3837, 2010.

[163] H. Zhang. The optimality of Naive Bayes. In *Proc. of FLAIRS Conference*, 2004.

[164] N. Zhou and L. Wang. Effective selection of informative SNPs and classification on the HapMap genotype data. *BMC Bioinformatics*, 8:484, 2007.

[165] M. Zuckermann, D.T. Larrance, J.A. Hall, R.S. DeFrank, and R. Rosenthal. Posed and spontaneous communication of emotion via facial and vocal cues. *Personality*, 47:712–733, 1978.

Own Publications

[166] **M. Karg**, R. Jenke, K. Kühnlenz, and M. Buss. A two-fold PCA-approach for inter-individual recognition of emotions in natural walking. In *Proc. of Int. Conf. Machine Learning and Data Mining*, 2009.

[167] **M. Karg**, R. Jenke, W. Seiberl, K. Kühnlenz, A. Schwirtz, and M. Buss. A comparison of PCA, KPCA and LDA for feature extraction to recognize affect in gait patterns. In *Proc. of Int. Conf. Affective Computing and Intelligent Interaction (ACII)*, pages 195–200, 2009.

[168] **M. Karg**, K. Kühnlenz, and M. Buss. A dynamic model and system-theoretic analysis of affect based on a piecewise linear system. In *Proc. IEEE Int. Symp. Robot and Human Interactive Communication*, 2009.

[169] **M. Karg**, K. Kühnlenz, and M. Buss. Towards a system-theoretic model for transition of affect. In *Proc. of AISB, Symposium on Mental States, Emotions and their Embodiment*, 2009.

[170] **M. Karg**, K. Kühnlenz, and M. Buss. Recognition of affect based on gait patterns. *IEEE Trans. Systems, Man, and Cybernetics - Part B: Cybernetics*, 40(4):1050–1061, Aug 2010.

[171] **M. Karg**, K. Kühnlenz, M. Buss, W. Seiberl, F. Tusker, and A. Schwirtz. Recognition of human states in natural walking. In *Proc. 1st Int. Workshop Cognition for Technical Systems*, 2008.

[172] **M. Karg**, M. Schwimmbeck, K. Kühnlenz, and M. Buss. Towards mapping emotive gait patterns from human to robot. In *Proc. IEEE Int. Symp. Robot and Human Interactive Communication*, pages 278–283, 2010.

[173] **M. Karg**, W. Seiberl, K. Kühnlenz, F.Tusker, M.Schmeelk, M. Buss, and A.Schwirtz. Expression and automatic recognition of exhaustion in natural walking. In *Proc. of IADIS Int. Conf. Interfaces and Human Computer Interaction (IHCI)*, pages 165–172, 2008.

[174] **M. Karg**, S. Sosnowski, K. Kühnlenz, M. Buss, C. Wendt, M. Popp, and B. Färber. Human emotion model using markov chains. In *Proc. 1st Int. Workshop Cognition for Technical Systems*, 2008.

[175] W. Seiberl, **M. Karg**, K. Kühnlenz, M. Buss, and A. Schwirtz. Analysis of human motion with methods from machine learning. In *Proc. of 27th Congress of the International Society of Biomechanics in Sports*, page 759, 2009.

[176] C. Wendt, **M. Karg**, K. Kühnlenz, and M. Popp. Emotion in human-robot interaction: Recognition and display. In *Int. Congress of Psychology (ICP)*, 2008.

[177] C. Wendt, M. Popp, **M. Karg**, and K. Kühnlenz. Physiology and HRI: Recognition of over- and underchallenge. In *Proc. IEEE Int. Symp. Robot and Human Interactive Communication*, pages 448–452, 2008.

Supervised Student Projects

[178] L. Cao. Feature extraction and classification of physiological signals. Master thesis, Lehrstuhl für Regelungs- und Steuerungstechnik, Technische Universität München, 2009.

[179] R. Jenke. Feature extraction for automatic recognition of emotions in gait patterns. Studienarbeit, Lehrstuhl für Regelungs- und Steuerungstechnik, Technische Universität München, 2009.

[180] R. Jenke. Analysis of the relation between Bayesian classification and statistical testing. Diplomarbeit, Lehrstuhl für Regelungs- und Steuerungstechnik, Technische Universität München, 2010.

[181] B. Karg. Ganganalyse mit SPSS. Praktikum, Lehrstuhl für Regelungs- und Steuerungstechnik, Technische Universität München, 2009.

[182] T. Schindl. Animation ausdrucksvoller Gangmuster. Studienarbeit, Lehrstuhl für Regelungs- und Steuerungstechnik, Technische Universität München, 2008.

[183] T. Schindl. Konstruktion von Animationen in C4D. Praktikum, Lehrstuhl für Regelungs- und Steuerungstechnik, Technische Universität München, 2009.

[184] M. Schwimmbeck. Gangalgorithmus mit variablen Gangmustern für den Hexapod SAM. Bachelorarbeit, Lehrstuhl für Regelungs- und Steuerungstechnik, Technische Universität München, 2009.

[185] C. Spies. Animation. Ingenieurspraxis, Lehrstuhl für Regelungs- und Steuerungstechnik, Technische Universität München, 2010.

[186] S. Toprak. Analyse des Blickverhaltens bei der Betrachtung emotionaler Gangmuster. Bachelorarbeit, Lehrstuhl für Regelungs- und Steuerungstechnik, Technische Universität München, 2010.

[187] X. Wang. Gait identification with regard to emotional states. Master thesis, Lehrstuhl für Regelungs- und Steuerungstechnik, Technische Universität München, 2009.