Karlsruhe Institute of Technology

Facial Expression Recognition using Surface Electromyography

Diplomarbeit am Cognitive Systems Lab
Prof. Dr.-Ing. Tanja Schultz
Fakultät für Informatik
von

cand. inform.
Martin Pruzinec

Betreuerin:

Prof. Dr.-Ing. Tanja Schultz

Tag der Anmeldung: 1. Oktober 2009
Tag der Abgabe: 29. Januar 2010

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Karlsruhe, den 29. Januar 2010

Abstract

Facial expressions are an essential part of human communication. The integration of information transferred by these expressions into the human-computer interaction can lead to affective systems that are able to adapt their behavior according to reactions shown by the user.

In this thesis we propose a system to detect and classify facial expressions based on surface electromyography (sEMG) signals. We recorded 4 sessions with each 50 samples of 7 posed facial expressions, including sadness, anger, disgust, fear, happiness, neutral and surprise. Different methods to process and extract signal information were investigated. For facial expression recognition we employed a Support Vector Machine (SVM) and achieved an average recognition rate of 92.02% session dependent and 86.79% session independent task. Furthermore, we reduced the electrode placement to 4 positions resulting in a mean classification rate of 87.86%.

In order to demonstrate the capabilities of our recognition system we built a live demonstration system in which facial expressions shown by a human are recognized and then mirrored by a 3D avatar.

Zusammenfassung

Der Forschungsbereich der Mensch-Maschinen-Interaktion (Human-Computer Interaction HCI) beschäftigt sich mit der Entwicklung von Schnittstellen zwischen menschlichem Benutzer und Maschine. Ein Hauptaugenmerk liegt dabei auf Systemen, die in der Lage sind, die natürlichen Kommunikationsweisen des Menschen zu erfassen und zu interpretieren. Unter den vielfältigen Möglichkeiten der menschlichen Kommunikation kommt neben der Sprache besonders Gesichtsausdrücken eine wichtige Rolle zu (z.B. das signalisieren von freundlicher Gesinnung durch sichtbares Lächeln). Der Begriff Gesichtsausdrücke umfasst in diesem Zusammenhang alle visuell erkennbaren Veränderungen im Gesicht, die auf kurze Zeitspannen begrenzt sind und eine Kommunikationsinformation enthalten.

In einer zwischenmenschlichen Konversation können Gesichtsausdrücke Zusatzinformationen zu Gesprochenem beinhalten (z.B. das verdeutlichen von Ironie durch Augenzwinkern oder Lächeln oder Unverständnis durch einen fragenden Blick des Zuhörers). Bereits 1872 beschrieb Darwin [Dar72] eine weitere Funktion von Gesichtsausdrücken. Er proklamierte die Existenz universell vorhandener Gesichtsausdrücke, welche bestimmte emotionale Stimmungen ausdrücken können. Dies wurde 1971 von Ekman und Friesen bestätigt [EF71]. Sie bestimmten 6 Gesichtsausdrücke, deren Bedeutung unabhängig vom kulturellen Hintergrund der Menschen erkannt werden. Die automatisierte Erfassung und Interpretation solcher Informationen könnte von Maschinen dazu genutzt werden, das eigene Agieren an einen Benutzer anzupassen, durch die als positiv oder negativ ausgewerteten Gesichtsausdrücke des menschlichen Benutzers.

Zur automatisierten Erfassung und Interpretation von Gesichtsausdrücken existieren zwei Hauptansätze, die Analyse visueller Informationen auf dem Gebiet der Computervisualistik sowie die Interpretation menschlicher Biosignale (z.B. durch Elektromyographie (EMG) erfasste Muskelaktivität). Der Einsatz von visuellen Techniken besitzt den Nachteil, dass diese in der Regel einen frontalen Blickwinkel auf das Gesicht benötigen und anfällig gegenüber Änderungen der Umgebungsbeleuchtung sind. Deshalb können solche Erfassungssysteme nicht in handliche mobile Geräte integriert werden. Im Gegensatz dazu kann EMG in kleinen mobilen Geräten ähnlich Kopfhörern oder einseitigen Headsets eingebaut werden.

In meiner Diplomarbeit beschäftige ich mich mit der Entwicklung eines Systems zur automatisierten Erkennung von Gesichtsbewegungen basierend auf EMG Signalen. Hierzu entwickelte ich zunächst ein Experiment zur Aufnahme von sieben Gesichtsausdrücken, welche die sechs Grundausdrücke Wut, Ekel, Angst, Trauer, Freude, Überraschung und einen entspannt neutralen Ausdruck beinhalten. Zur Aufnahme der Muskelaktivität wurden acht Oberflächen-Elektrodenpaare über den mimischen Hauptmuskeln platziert. Mit diesem Aufbau erfasste ich in vier Aufnahmesitzungen jeweils 50 posierte Darstellungen pro Gesichtsausdruck. Hierfür wurden drei männliche Probanden erfasst, wobei ein Proband zwei Aufnahmesitzungen an einem Tag mit einer erneuten Platzierung der Elektroden zwischen den Aufnahmesitzungen absolvierte.

Im Verlauf der Diplomarbeit entwickelte ich ein System zur Extraktion und Klassifikation von Merkmalen aus dem EMG Signal. Dabei untersuchte ich verschiedene Methoden zur Verarbeitung und Extraktion von Merkmalen des EMG Signals und verwendete zur Erkennung der Gesichtsausdrücke Support Vector Maschinen (SVM).

Für die Auswertung des Systems wurden alle Aufnahmesitzungen für jeden Gesichtsausdruck in 30 Trainings- und 20 Testbeispiele aufgeteilt. Zunächst untersuchte ich die Erkennungsraten verschiedener Kombinationen von Parametern und Extraktionsmethoden. Pro Aufnahmesitzung eines Probanden wurden bei Verwendung der zugehörigen Trainings- und Testbeispiele im Mittel eine Erkennungsleistung von 92.02% erreicht. Beim sitzungsunabhängigen Test auf Daten des gleichen Probanden ergab sich im Mittel eine Erkennungsleistung von 87.79%. Des weiteren untersuchte ich eine systematische Reduzierung der Elektrodenanzahl und erreichte eine Reduktion um 50% auf vier Positionen mit einer mittleren Erkennungsrate von 87.86%. Dieser reduzierte Elektrodensatz umfasste keine Positionen unter den Augen, nahe der Nase oder des Mundes. Dies führt dazu, dass keine Kabel durch das Gesicht geführt werden müssen. Somit wäre auch eine Integration in ein Alltagsgerät möglich (zB. ein Headset für mobile Kommunikation).
Um einen potentiellen Einsatz in der maschinenbasierten Mensch zu Mensch Kommunkation zu demonstrieren, entwickelte ich abschließend ein Echtzeit-System, bei dem die Gesichstbewegungen eines menschlichen Nutzers erkannt und auf einem animierten 3D Kopf gespiegelt dargestellt werden.

Contents

1 Introduction **1**
- 1.1 Motivation . . . 1
- 1.2 Goal . . . 2
- 1.3 Structure of this Thesis . . . 2

2 Facial Expressions and Electrophysiological Background **3**
- 2.1 Facial Expressions . . . 3
  - 2.1.1 Measuring Facial Expressions . . . 4
- 2.2 Physiological Fundamentals . . . 5
  - 2.2.1 Skeletal Muscles Structure . . . 5
  - 2.2.2 Muscle Contraction . . . 5
- 2.3 Facial Muscles and Electrode Placement . . . 7
  - 2.3.1 Facial Muscles . . . 7
  - 2.3.2 Electrode Placement . . . 9
  - 2.3.3 Anatomical Variation and Facial Expression . . . 9
- 2.4 Electromyography . . . 10
  - 2.4.1 Surface Electromyographic Measurement . . . 10
  - 2.4.2 Equipment . . . 10
- 2.5 Characteristics of EMG . . . 11
  - 2.5.1 Signal Quality Dependencies . . . 12
  - 2.5.2 Signal Distortions . . . 12

3 Related Work **13**

4 Data Collection **15**
- 4.1 Expression Set . . . 15
- 4.2 EMG Equipment . . . 15
- 4.3 Electrode Placement . . . 16
  - 4.3.1 Estimated Activation Pattern . . . 17
- 4.4 Recording Setup . . . 20
  - 4.4.1 General Setup . . . 20
  - 4.4.2 Subject Introduction . . . 20
  - 4.4.3 Stimuli Presentation . . . 21
  - 4.4.4 Software . . . 22
- 4.5 Data Corpus . . . 22

5 Methods **23**
- 5.1 Offset Substraction . . . 24
- 5.2 Feature Extraction . . . 24

5.2.1 MimikFeat . . . 24
5.2.2 TD . . . 26
5.3 Feature Normalization and Scaling . . . 28
5.3.1 Normalization . . . 28
5.3.2 Scaling . . . 28
5.4 Feature Reduction . . . 28
5.5 Electrode Reduction . . . 29
5.6 Support Vector Machines . . . 29
5.6.1 Basic Principle . . . 30
5.6.2 Soft Margin . . . 31
5.6.3 Multi-Class Classification . . . 31
5.6.4 Radial Basis Function . . . 32
5.6.5 Parameter Selection . . . 32

6 Evaluation **33**
6.1 Expression Recognition . . . 33
6.1.1 System Structure Description . . . 33
6.1.2 Feature Set Selection . . . 35
6.1.3 Window Size and Context Length . . . 36
6.1.4 Specific Expression Recognition Characteristics . . . 37
6.1.5 SVM Parameter Optimization . . . 38
6.1.6 Session and Subject Independent Classification . . . 40
6.1.7 Reduction of Electrode Positions . . . 41
6.1.8 Conclusions . . . 43
6.2 Live Expression Recognition Demonstration . . . 43
6.2.1 Thinking Head . . . 43
6.2.2 Simulation System . . . 44
6.2.3 Live Demonstration System . . . 46

7 Conclusion **47**

A Evaluation Result Tables **49**

Bibliography **55**

List of Figures

2.1 Structure of a skeletal muscle. (Original by [Sch06]). 6

2.2 Facial muscles: higher muscle layers on the right and deeper muscle layers on the left side [Sob00]. 8

2.3 Electrode positions for recording the main facial muscles by Cacioppo et. al [FC86] 9

2.4 The bipolar measurement of action potentials with surface electrodes. (Original by [LH92]) 11

4.1 Electrode placement with channel numbers: M. Frontalis (1), M. Corrungator (2), M. Orbicularis Oris (3), M. Levator Labii (4), M. Zygomaticus (5), Masseter (6), M. Depressor Anguli Oris (7), Mentalis (8) (Background Image, [SSSbKW06]) 16

4.2 Probabilistic channel activation of mapped data from [GKF95]. a) Relative Mean b) Relative Variance. (1) M. Frontalis (2) M. Corrungator (3) M. Orbicularis Oris (4) M. Levator Labii (5) M. Zygomaticus (6) Masseter (7) M. Depressor Anguli Oris (8) Mentalis 19

4.3 Probabilistic channel activation of mapped data from [GKF95] for each of the eight channel ordered by expression. 19

4.4 Stimuli for the seven expressions: Anger, Disgust, Fear, Happiness, Neutral, Sadness and Surprise (Created with [SF09]) 21

4.5 Presentation Cycle: Focusing, Stimulus Presentation and Relaxation. The Sync Channel marks if a picture is displayed. 22

5.1 General processing chain from raw sEMG to SVM model training. . . 23

5.2 Example of an SVM hyperplane with maximum margin trained on data samples of two classes. 30

6.1 General structure diagram of the evaluation system. 34

6.2 Average expression recognition rate (ERR) of TD0, MimikFeat1 and MimikFeat2 on different window sizes. 35

6.3 Average single feature vector recognition rate VRR of TD0, MimikFeat1 and MimikFeat2 on different window sizes. 36

6.4 Average ERR and VRR of TD0, TD-Slope, TD-Waveform, TD-Slope-Waveform and MimikFeat2 on different window sizes. 37

6.5 SVM parameter optimization structure. 39

6.6 Training phase of the SVM model 44

6.7 Demonstration phase: The analog EMG signal is analyzed for channel activation and the expression is recognized. The recognition result is sent to a TH plugin and the expression is emoted on the TH 3D head. 45

List of Tables

2.1 The major facial mimic muscles and their functions. 7

4.1 Specifications of VarioPort III 16

4.2 Mapping of FACS Action Units onto the Electrode Placement. 17

4.3 Definition of Action Units (AUs) in the Facial Action Coding System [ER05] 18

6.1 Average confusion matrix over the test sets of all sessions, with TD0-Slope-Waveform and window size 400 ms. 38

6.2 SVM parameter optimization results for the recorded sessions, with TD0-Slope-Waveform 40

6.3 ERR results for session independent training and testing. 41

6.4 SBCR results on a session combined SVM model and ERRs for single session SVMs, trained and tested on the channel subset. 41

6.5 Average confusion matrix of channel subset [1 2 5 6]. 42

A.1 Expression Sample Recognition Rates (EER) for TD0, MimikFeat1, MimikFeat2 50

A.2 Feature Vector Based Recognition Rate (VRR) for TD0, MimikFeat1, MimikFeat2 51

A.3 Average Majority Decision Expression Sample Recognition Rate (ERR) of TD Variants 52

A.4 Average Feature Vector Based Recognition Rate (VRR) of TD Variants 53

A.5 Time in ms to produce one TD Feature Vector in relation of different Context Length and Window Sizes 54

1. Introduction

1.1 Motivation

The research field of human-computer interaction (HCI) deals with the development of communication interfaces between human users and machines. One major interest is to create affective systems that are able to detect and interpret all natural ways of human communication.

In the many ways of direct human-human communication facial expressions pose an essential role. The information visible in the human face is transferred through different types of signals including static facial parameters and facial changes of different speeds. The most interesting changes are rapid motions, commonly known as facial expressions.

Facial expressions can transfer many different types of information, including additions to spoken messages, such as combining sentences with a wink to support a pleasant information. Furthermore, C. Darwin already stated in 1872 that facial expressions can express a current emotional state [Dar72], e.g. by displaying basic universal expressions like joy through a smile. Mehrabian [Meh68] states that to communicate emotions in human-human interaction about 55% of the information is signaled through facial expressions, while only about 7% is sent through spoken language and 38% by paralanguage, such as body gestures.

Including facial expression recognition into human-robotic communication can provide affective systems that react to the facial expression of a user. For example, a robot could gather negative or positive feedback for its behavior by interpreting the facial expressions of a user. Furthermore, dangerous or unexpected situations could be identified by the detection of expressions visualizing user reactions like fear or surprise. In the future this can enhance the security of machine control systems, e.g. in car control the systems for braking and stabilization can be prepared to provide faster reactions to a subsequent user input.

In this thesis we develop a facial expression recognition system based on facial surface electromyography (sEMG). The sEMG signal provides information about the muscle activity in a detection area. The sEMG is measured by attached surface electrodes.

The advantage of an sEMG based approach in expression recognition, compared to a video based approach, is the possibility to develop a small mobile device and therefore generate a lightweight system that can detect facial movements.

Furthermore, a system that is able to detect and interpret facial muscle activities can be used directly as an input device. A possible application scenario in the field of medicine is to enable paralyzed patients to control technical equipment and provide additional possibilities in communication with their environment by using voluntarily mimic muscle changes as control commands. Especially for quadriplegic patients[1] this would provide a strong enhancement of their independency.

1.2 Goal

The main goal of this thesis is to develop a system that is capable to detect and classify a set of basic facial expressions based on activity patterns of mimic muscles measured with surface electromyography (sEMG). This system will provide a way to use voluntary facial expressions as an alternative input to machine systems. There is no standardized database of sEMG facial expressions, therefore a data corpus has to be developed.
We investigate different methods to process and extract signal information. Furthermore, we do research on the generalization capability of the developed system, when trained and tested on different recording sessions and persons. Furthermore, we investigate a minimization of the amount of electrode positions, in order to decrease the distraction of a human user and determine a practicable set of electrode positions.

Finally, we build an online demonstration system that is capable to classify a posed facial expression and displays the result on a 3D avatar to demonstrate a possible communication scenario.

1.3 Structure of this Thesis

This thesis is structured into 7 chapters. In the current chapter we describe the motivation and goals of this thesis. In chapter 2 we introduce the topic of facial expressions and their physiological background. Furthermore, we describe the way surface electromyography is detected and can be applied to measure mimic muscle activity. In chapter 3 work related to sEMG based facial expression recognition is presented. Chapter 4 describes the set of expressions and electrode positions and the experimental setup.
In chapter 5 the general system structure and methods are introduced. Followed by the evaluation in chapter 6, where we determine a classification system that includes the best combination of these methods and parameters in regard to the expression classification performance. In conclusion, chapter 7 summarizes this thesis and suggests directions for future research.

[1] Patients that are paralyzed at both arms and both legs, caused by a high spinal cord accident.

2. Facial Expressions and Electrophysiological Background

In this chapter the principles of facial expressions and the physiological basis of their creation are described. Furthermore, the basics of surface electromyography (sEMG) and the way it can be used to measure facial expressions are introduced. Concluding characteristics of EMG signals are described.

2.1 Facial Expressions

The human face forms a central part in verbal and nonverbal human-human communication. Nonverbal communication includes many different types of signals, such as body movements, posture, gaze, voice modulation and information displayed on the face [CE05].

The face provides many kinds of information that are transferred by different types of *sign vehicles*, which are meaningful facial cues. These can be divided into four general groups according to P. Ekman [Ekm78]: *Static vehicles* represent permanent features, for example the bone structure can display ancestry. *Slow sign vehicles* are changes in the appearance of the face over time, e.g. permanent furrows can demonstrate a person has spent much time in the open. *Artificial sign vehicles* represent added artificial facial features, such as eyeglasses and cosmetics. The most prevalent are *rapid sign vehicles* that represent detectable visual changes in the facial appearance [HE95]. These rapid changes are produced by deformations of the facial skin and mouth opening caused by muscle movements.

In this thesis we are interested in rapid sign vehicles, because they include the information that can voluntarily be changed in communication. These vehicles can express specific meanings, e.g. smiling displays a joyful perception. From now on we refer to these rapid changes as *facial expressions*.

A systematic study of facial expressions and their causes is "Essays on the Anatomy of Expression in Painting" and was published in 1806 by Sir Charles Bell. He was

the first scientist to determine the existing facial muscles in detail. He assumed that these muscles have the only purpose to display expressions. In 1862 Duchene de Boulogne performed a expreiment on a subject, which did not feel any pain. He induced electrical currents to the face of the subject studying the resulting transformations. He reported which muscles act to produce some of the facial expressions.

In human-human communication many communicative signs are transferred through facial expressions. These signals can accompany spoken conversations as *syntactic displays* to stress words and clauses or as *speaker displays*, to illustrate the idea conveyed. Furthermore, they can visualize a response to an utterance as *listener comment displays* [NT94].

Charles Darwin claimed in "The Expression of Emotion in Man and Animals" [Dar98] the existence of universal expressions displaying emotional states in humans and some animals.
This work of Darwin then had not been recognized in science for about a hundred years [Dar98], until these findings were confirmed by Ekman et al. [EF71] in 1971, describing primary universal expressions. These are reflecting distinguishable emotions and are recognized from humans independently of their cultural background. These expressions are also referred to as the *basic emotions*, they include anger, disgust, fear, happiness, sadness and surprise.

2.1.1 Measuring Facial Expressions

Two major approaches exist to measure facial expressions: visually investigating visible facial changes or analyzing physiological signals. Thereby, the rapid signal vehicles are manually coded by humans or automatically processed by computer systems.

For the manual coding, systematic description systems have been developed to measure rapid facial changes by describing a facial expression as a combination of small movements. The most commonly used system is the Facial Action Coding System (FACS) introduced by Ekman and Friesen [EF78] in 1978.

FACS is an anatomical based system to measure visually distinguishable facial movements [ER05]. In the developing process of this system, Ekman and Friesen used needle EMG to verify that they included all possible facial movements into FACS. Facial expressions are described by a combination of 44 unique action units (AUs). Additional categories are included for head and eye positions and movements. Thereby, the AUs do not reflect single muscles, but describe observable changes in facial areas. Some AUs are produced by the same muscle through different activation patterns, for example AU 1, the "Inner Brow Raise", and AU 2, the "Outer Brow Raiser", refer to movements caused by the contraction of different parts of the musculus frontalis.

In the last few years, considerable effort has been made toward an automatic measurement and recognition of facial expressions through computer vision. Here, features are extracted from digitized images that represent meaningful patterns, like directions and intensity of motions. These features are then used for recognizing different expressions [CE05]. A survey into current approaches of visual expression classification is given by Fasel et al. [FL03].

A facial expression is a rapid visual skin movement caused by the activity of facial muscles. Two main approaches exist to describe facial expressions by physiological

signals, henceforth referred to as *biosignals*. Facial surface EMG directly measures muscle activity at different positions in the face. EMG detects differences in electrical potentials through electrodes indicating the level of muscle contractions. The advantage of this method is the high temporal resolution [CE05]. Small muscle activities can be detected that do not result in visual detectable changes. A more detailed description is given in section 2.4. A second method is electroencephalography (EEG). It can be used to measure activation patterns of the motor cortex in the human brain preceding a muscle movement. The disadvantages of this method, compared to EMG, are the relatively low voltage signals and high susceptibility to EMG signals and electromagnetic noise.

2.2 Physiological Fundamentals

The facial muscles involved in the production of facial expressions are known as *mimic muscles*. The mimic muscles are mainly *skeletal muscles* and consist of bundles of parallel *muscle fibers*. Each of these muscle fibers is directly activated by one *motor neuron* [DSH05]. The motor neurons themselves are stimulated through the nervous system. The motor neuron, its axon and all of its innervated muscle fibers are called *motor unit* [And07]. The following sections will give an introduction into the structure and function of skeletal muscles.

2.2.1 Skeletal Muscles Structure

Generally, skeletal muscles are attached to bones and move the skeleton. In the case of facial muscles they are attached to bones and skin. Their function is restricted to moving the facial skin. An exception is the *masseter*, which is responsible for mouth opening by moving the jawbone.
Figure 2.1 shows the principle structure of a skeletal muscle. Each skeletal muscle consists of several bundles of muscle fibers. The muscle fibers themselves consist of several *myofibrils* that include contractable sections of *sarcomeres*. These sections include *myofilaments* of actin and myosin and several further structure and cytoskeletal proteins. The actin and myosin filaments are arranged parallelly to each other in the sarcomere. Skeletal muscles are innervated through a neuromuscular junction, also known as the motor endplate. It is located near the middle of each muscle fiber. For a more detailed description of the structure and chemical processes refer to Deentjen [DSH05] and Schandry [Sch06].

2.2.2 Muscle Contraction

Muscle contractions are controlled through the nervous system. Before a muscle contracts, a control command is sent from the spinal cord or the motor cortex. The *motor cortex* is a region in the cerebral cortex of the human brain, which is involved in the planning and control of muscle movements. The control command is send in form of an *action potential* (AP), an electrical impulse, running along the axon of a motor neuron towards connected muscle fibers. The *axon* is a long threadlike part of a nerve cell that conducts impulses to another cell. When the AP reaches a motor endplate, the neurotransmitter *acetylcholine* (ACh) is released. In this way the electrical stimulation is transmitted chemically to the muscle fiber following the

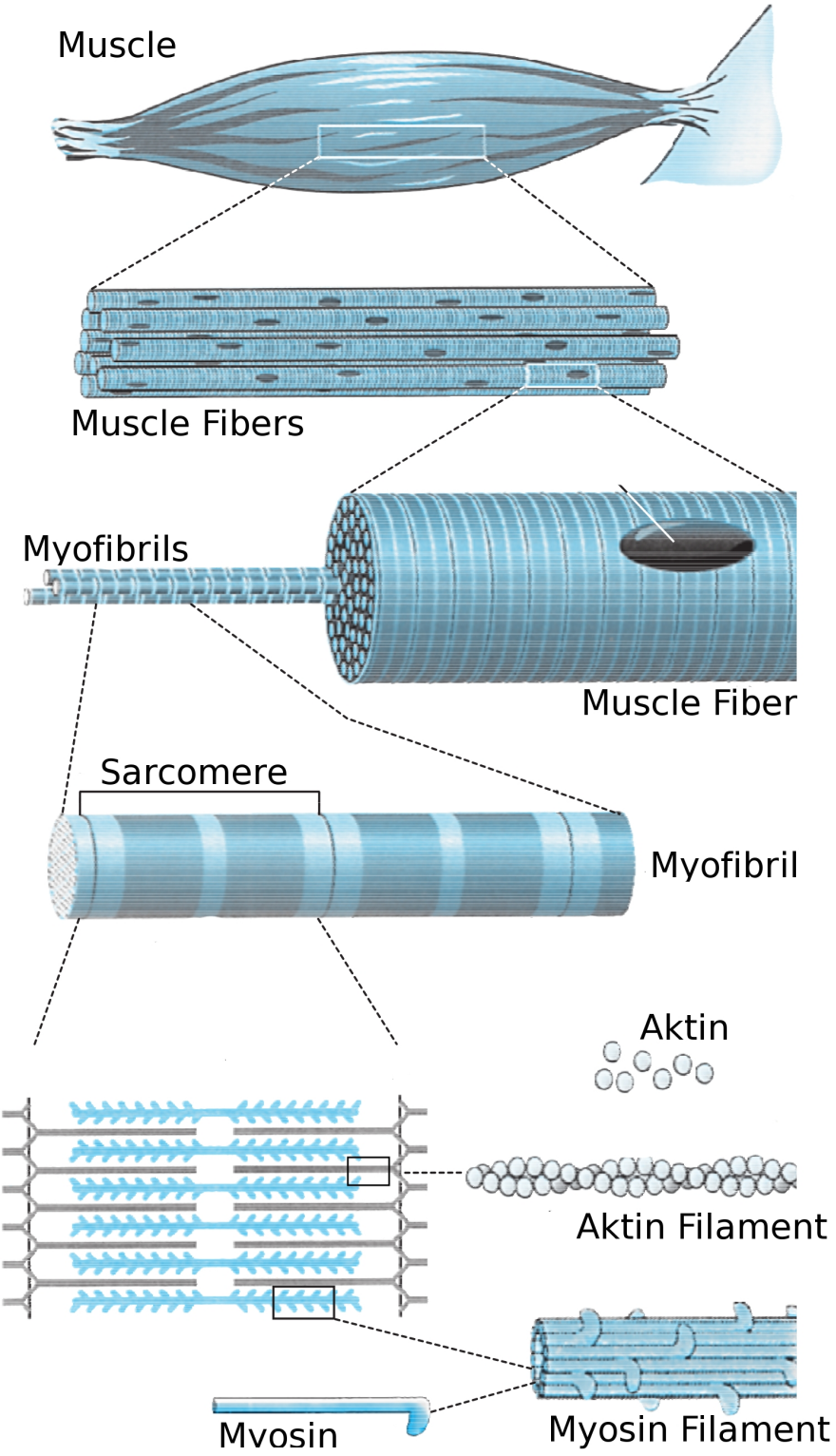

Figure 2.1: Structure of a skeletal muscle. (Original by [Sch06]).

all-or-none law. The all-or-none law says that an activation is transferred only if it reaches a certain level and then the activation is maximal.

Therefore, a certain amount of ACh leads to a depolarization of the muscle fiber surface. This creates an AP that travels towards both ends of the muscle fiber. This AP leads to the release of calcium ions into the sarcoplasm surrounding the myofibrils. This causes myosin to change its structure and pull on actin filaments, which then is sliding between the myosin spaces. In this way the muscle contracts. The power and scale of muscle contraction is given by the the amount of simultaneously contracting muscle fibers and the frequency of APs.

The electrical action potentials of muscle fibers can be detected through EMG. The level of detection detail of a potential is strongly dependent on the type of EMG equipment and sensor placement. In the section 2.4 we give an overview of the functionality and characteristics of EMG.

2.3 Facial Muscles and Electrode Placement

2.3.1 Facial Muscles

There are about 40 mimic muscles in the human face. Their major function is to move facial skin. Facial muscles are mainly innervated by the *nervus facialis*. The jaw muscles, innervated by nervus trigeminus, also move the jaw bone, e.g. masseter. The nerves are bilaterally organized. In this way, each facial side is innervated by a different nerve. The movements of both facial sides origin from a separate motor nucleus in the human brain and are connected by direct neuronal links. Therefore muscles on both facial sides are separately controlled and symmetrical motion has to be timed in the motor nerve system [Rin84].

Facial muscles are controlled by a relatively large area on the motor cortex. The upper facial muscles are represented by a smaller area than the lower muscles. This explains the existence of higher motor skills in the lower face than in the upper part. For example, most people can voluntarily move one side of the mouth but only few can move the eyebrows separately.

M. Frontalis:	Raises the eyebrow and wrinkles the forehead skin.
M. Corrungator Supercilii:	Pulls the eyebrow to the medial corner and down, causes vertical wrinkles between the eyebrows.
M. Orbicularis Oculi:	Raises the cheek and narrows the eye.
M. Levator Labii Superioris:	Raises upper lip, wrinkles the nose.
M. Zygomaticus Major and Minor:	Pulls mouth corner lateral upwards.
M. Orbicularis Oris:	Controls mouth shape and size.
Masseter:	Presses jaws together. Mouth opens when released.
M. Depressor Anguli Oris:	Lowers mouth corner.
Mentalis:	Pushes chin up and wrinkles chin boss.

Table 2.1: The major facial mimic muscles and their functions.

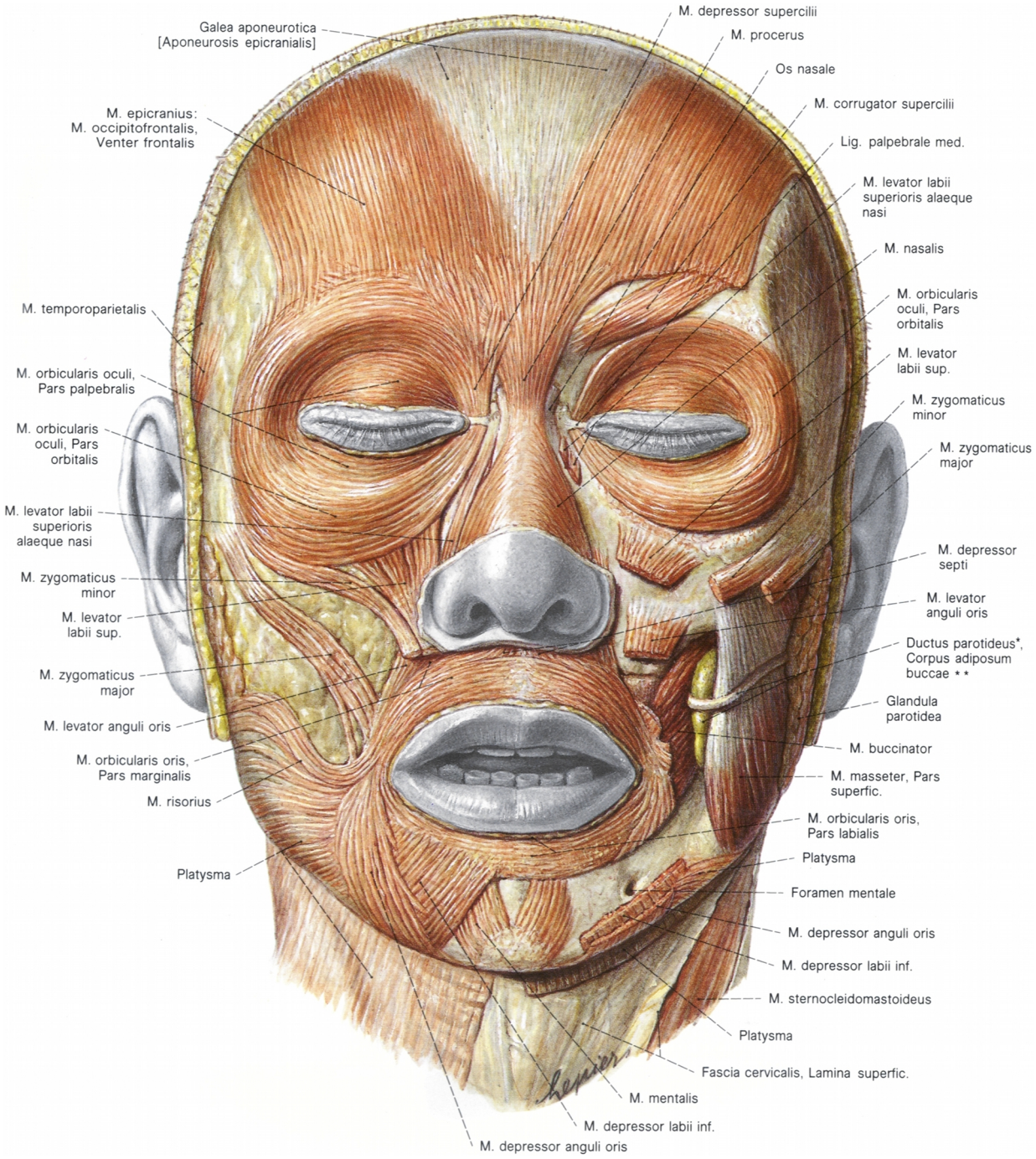

Figure 2.2: Facial muscles: higher muscle layers on the right and deeper muscle layers on the left side [Sob00].

Within the mimic muscles, there is a large variance in shape, size, muscle type and function. The general functions of the main facial muscles are listed in Table 2.1. Furthermore, in the human face the density and crossing of muscles in different layers change widely (see Fig. 2.2). More muscles overlap in the middle facial region close to the eyes, mouth and nose.

2.3.2 Electrode Placement

To record facial expressions with surface EMG the electrode placement has to cover the major mimic muscles. Therefore, the muscles that are supposed to strongly participate in facial expressions have to be included. The electrode placement positions can be described in relation to anatomical landmarks. This provides the possibility to identify electrode positions in different faces.

In 1986 Cacioppo and Fridlund [FC86] published placement guidelines to support comparability for facial EMG research. They describe appropriate electrode positions to measure the activity of the major facial muscles. The full placement set is shown in Figure 2.3.

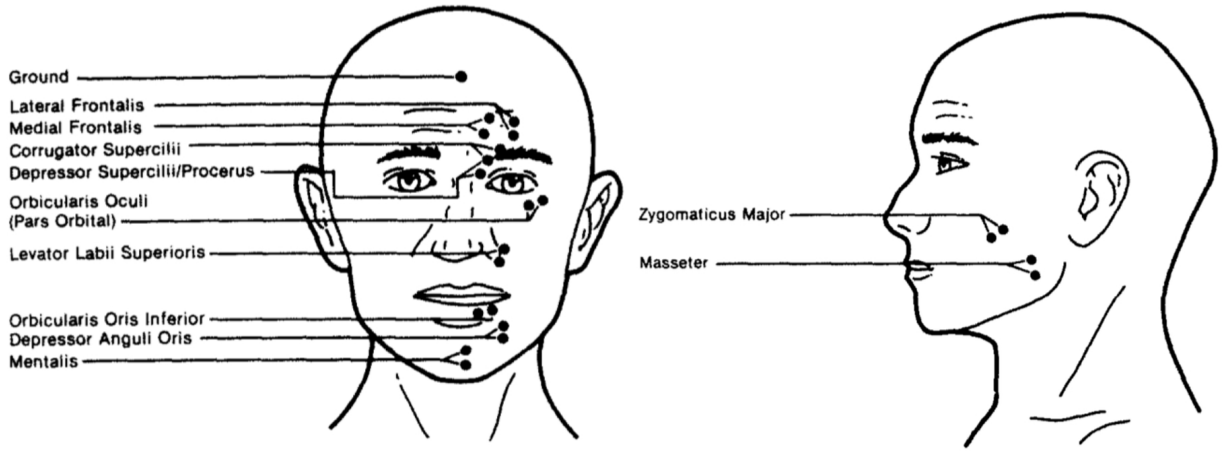

Figure 2.3: Electrode positions for recording the main facial muscles by Cacioppo et. al [FC86]

2.3.3 Anatomical Variation and Facial Expression

Individual differences exist in the structure of facial muscles (see Schmidt and Cohn [SC01]). Muscle bundles like the risorius, platysma and zygomaticus major are highly variable between humans. For example, in most individuals, the platysma muscle inserts at the skin above the inferior margin of the mandible. But it was also observed to insert the lateral cheek and thereby causing a vertical depression or wrinkles to appear. The m. zygomaticus major varies by appearing also in a split version with two separate insertion points. This appeared with 17 of 50 of subjects in an anatomical study by Pessa et al. [PZA+98]. When smiling, the tension by the two muscle parts on the mouth corner causes a dimple during the muscle contraction. These anatomical variations in facial muscles contribute to individual differences in expression.

2.4 Electromyography

The measurement of electrical potentials during the contraction of muscle fibers is called *electromyogram*. These potentials are APs, which travel along the muscle fibers during a contraction phase. The electromyogram can be detected through electromyography. Hereby, the time and spatial resolution of the measured signal strongly depend on the method and EMG equipment used.

For example, glass micro needle electrodes can be directly introduced into a muscle fiber. This method is able to precisely detect muscle activations on a level of single muscle fibers. The biggest drawback of needle electrodes in general is that the placement requires medical knowledge and brings health risk by intrusion in the human body. Therefore, they are not applicable to be used in a human-computer interface outside medical environments or laboratories.

A non-invasive alternative is surface electromyography (sEMG). Here, electrodes are attached to the body surface over a muscle of interest. The sEMG only detects a portion of the existing action potentials that is transmitted through the skin. On the one hand, the muscle fibers surrounding an active fiber are functioning as an electrical insulation and on the other hand, the fat and skin tissues are functioning as a lowpass frequency filter. Therefore, "the closer the muscle tissue is to the skin surface and the stronger the contractions, the greater will be the amount of electrical activity occurring in large-muscle groups" [And07]. In this way, sEMG does not measure the activity of a specific muscle fiber, but the summation of activation potentials of several muscle fibers reaching the surface. The surface EMG is more likely to be a measurement of activities of the area next to the placed electrodes than a detailed measurement of single muscles contractions.

2.4.1 Surface Electromyographic Measurement

For sEMG measurement, electrodes are placed on the skin longitudinally to the length axis of a muscle of interest. In general one (unipolar) or two (bipolar) electrodes are placed at the recording site. The *unipolar* EMG signal is the potential difference between one electrode placed at the detection site and one or two reference electrodes placed on electrical neutral tissue. The signal from a *bipolar* setup is the difference in the measured potentials between the two electrodes placed at the detection site.

Figure 2.4 shows an exemplary bipolar measurement with two electrodes A and B, which are placed longitudinal to a muscle. An activation potential starting on the left site of A will reach this electrode before it reaches the region under B. The polarization of the differential signal gives information about the location of the highest muscle activation.

2.4.2 Equipment

For sEMG measurement electrodes are attached on the skin surface to pick up muscle activity. There are two different types of electrodes that can be employed: *dry*

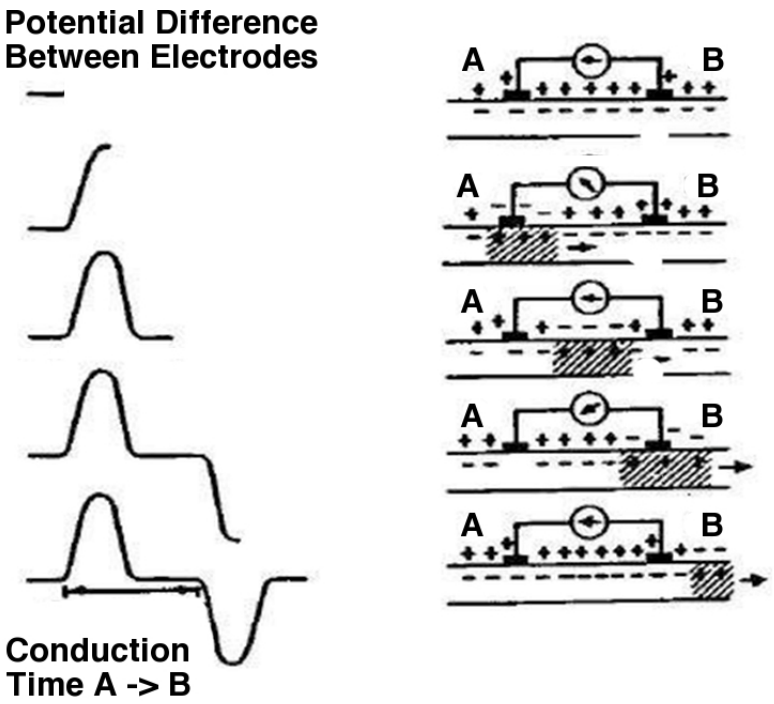

Figure 2.4: The bipolar measurement of action potentials with surface electrodes. (Original by [LH92])

electrodes that have direct skin-contact and *gelled electrodes*, which use an electrolytic gel as a chemical interface between the metallic electrode part and the skin to decrease the skin-electrode impedance [Sco03].
When detecting the EMG signal, amplification is necessary. To optimize the resolution of digitizing equipment and to minimize noise the amplification should be applied close to the electrode [Sco03]. Therefore, the electrodes are commonly connected via thin wires to an amplifier.

The analog signal usually has to be digitized by an analog to digital (A/D) converter for further processing and analyzing. Furthermore, for data storage a direct connection to a data processing device is necessary, e.g. a serial interface connected to a computer. A detailed description of the equipment used in this thesis is given in chapter 4.2.

2.5 Characteristics of EMG

The signal quality recorded by facial sEMG is determined by the used equipment, the environmental surrounding and the electrode placement. The frequencies of major signal activity range between 20 - 250 Hz [DL02]. In the frequency segment lower than 20 Hz mainly movement artifacts are located. The amplitudes of the measured signal ranges between 0.1 - 5 mV. However, higher values can be reached depending on the tissue mass under the electrodes and movements through contraction [And07] [DL02].

2.5.1 Signal Quality Dependencies

The sEMG signal is influenced by many factors including physiological, anatomical and technical issues. According to [Sco03] the amplitude, time and frequency domain properties of the sEMG signal depend on the following:

- the timing and intensity of muscle contraction,
- the distance of the electrode from the active muscle area,
- the properties of the overlying tissue (e.g. thickness of overlying skin),
- the electrode and amplifier properties, and
- the quality of contact between the electrode and the skin.

Usually information about the time and intensity of muscle contraction is desired. These five factors only increase the variability in the sEMG recordings. The impact of these non-muscular factors can be minimized by using the same equipment and electrode placement in different recording sessions.

2.5.2 Signal Distortions

The sEMG signal is further interfered by noise of different kinds. The noise consists of *inherent factors* in the electrical equipment and *external factors*. The inherent noise cannot be eliminated, but can be reduced by employing high quality components. The external factors can be categorized into four types [RHY06, DL02]:

- *Ambient noise*, which is electro-magnetic radiation e.g. through power lines and monitors at 50Hz or 60Hz. The body surface constantly absorbs these radiations and it is impossible to eliminate them. The peak of the ambient noise signal may be one to three orders of magnitude bigger in amplitude than the EMG Signal.
- *Motion artifacts* lower or raise the average signal from its baseline. They occur in the frequency range between 0 and 20 Hz and can result from disruptions of the skin-electrode interface or movements of cables connecting electrodes and amplifier.
- *Electrode contact* determines the signal-to-noise ratio. Therefore, careful preparation of the skin and electrodes is essential, e.g. applying alcohol to the skin before placement.
- *Crosstalk* is the intermixture of pure muscle activity signal with activity potentials of neighboring muscles. With sEMG the differentiation between crosstalk and the activity of one specific muscle is an almost impossible task. Therefore, it is more likely to record the activation of facial muscle areas than the activation of single muscles [FC86]. The crosstalk can be minimized by selecting electrode positions distant to areas with high density of different muscles.

3. Related Work

This chapter gives an overview of research approaches on facial surface electromyography. The major research fields are: the control of devices via facial muscle contractions, an evaluation of the emotional state of a person, facial expression detection and silent speech recognition based on surface EMG.

The general approach to detect and interpret facial surface EMG signals is to analyze the changes of the mean activation value at different electrode positions. Here, simple threshold methods are implemented and a recorded signal is interpreted into two classes of activation states ('on' or 'off'), e.g. mouse control via left and right side contraction of the m. temporalis [JKS05] and control of an robot hand by detecting activations of the left and right masseter [SK05].

Another research field in interpreting the facial EMG signal is the determination of the emotional state of a person. Various studies demonstrated that changes in the activity of certain facial muscles that occurre in reaction to strong stimuli can be linked to specific emotions [Dim90] [LNC03]. The surface electrodes are usually applied at the positions of the m. corrungator supercilii (CS) and the m. zygomaticus major (ZM). The signals detected at these positions show linear correlations to emotional responses, which are reactions related to emotions that follow on an affective stimulus, such as feeling happy when seeing a laughing child. An increase of the EMG signal activity at the CS is linearly correlated to a negative stimuli and at the ZM to a positive stimuli. These findings are used in applications like the evaluation of radio advertisement [BLP01]. Furthermore, this is used to evaluate the progresses of medical treatments by comparing the change in emotional responses to certain stimuli over time [CHOA81].

A detailed work on the classification of facial expressions representing an emotional state was done based on facial sEMG [FSF84]. The authors investigated the classification of four basic emotional expressions: sadness, anger, happiness and fear. From 12 female subjects 3 samples of posed expressions were recorded. They used 4 electrode sites (channels) including the m. frontalis, the m. corrugator supercilii, the m. orbicularis oculi and the m. orbicularis oris. To train and test a classification system the mean value of the EMG signal was estimated for each of the 4 channels

for every sample. Separately for each subject, a linear discriminative analysis was employed on this data.

An average recognition rate of 86% over all four expressions and subjects were achieved. The average recognition rate for single subjects ranged between 67% and 100%. The study employed the same data for training and testing as the classification system. Therefore, the recognition results are expectedly high because of a classification system optimized on the test data. They proved the general possibility of recognizing basic facial expressions with facial sEMG. The authors suggest that more electrode positions have to be recorded or more complex classification methods have to be employed to improve classification results on unseen data.

A different approach was made by [ABB+04], who analyzed the classification of 3 expressions: happiness, anger and sadness. EMG signals were measured at 3 positions at the masseter, the m. levator labii superioris and the m. corrugator supercilli. They recorded 6 subjects each performing 30 examples for each expression. Features of every example signal were extracted for each channel: the mean value, the standard deviation, the root mean square (RMS) and the power density spectrum deviation. From these 12 features the subset of 3 features was selected that provided the highest discrimination between classes. Then, templates of each expression class were calculated as the average over all 30 examples. The distance of the 3 dimensional vectors of an example expression to the three class templates was estimated and the minimal distant template then used as the predicted expression class. They achieved classification results between 84.44% and 100%. Again, this study did train and test on the same data.

An approach of a more detailed interpretation of facial EMG signals is done in the automatic silent speech recognition based on sEMG signals. The possibility to recognize a small number of words based on surface EMG was demonstrated by [CEHL02]. [MHMSW05] showed the possibility of speaker and session independent recognition of a small vocabulary. In [WS09] the approach of continuous silent speech recognition using adaptive speaker models was described.
To recognize silent speech using surface EMG, the signal measured at different facial electrode positions is split into two signal parts of mainly high and low frequencies. These signal parts are then sampled into small time segments. On these segments different features were extracted, including mean, zero crossing rate and frame based power. To include further context information about the current segment, a context filter was applied combining values of neighboring segments. A linear discriminative analysis was employed for dimension reduction. This data then was used to train and test Hidden Markov Models (HHMs). This method proved to achieve good classification results.

In this thesis we investigate the possibilities of facial expression classification by applying more electrode positions and more expressions than previous studies. We also analyze a larger set of examples and use separate training and test sets. We develop a system motivated by the methods used in continuous silent speech recognition.

4. Data Collection

A data collection had to be conducted prior to the experiments described in chapter 6. This chapter will give a description of the set of facial expressions we included in our experiments. Then, details on the used EMG equipment and the electrode placement are given. Subsequently, we describe the general experimental design. In conclusion, the detailed description of the data corpus is listed.

4.1 Expression Set

For our study, a set of basic facial expressions was recorded. In their research on facial expressions Ekman and Friesen [EF71] determined a set of six expressions that are universally recognized by humans independent from their cultural background (see section 2.1). These basic expressions are anger, disgust, fear, happiness, sadness and surprise. This set is also commonly referred to as the *basic emotions*, as those expressions can also display the general emotional state of a person. In the facial expression research this set of expressions is widely used. Some of these expressions do not necessarily contain strong facial activities when posed, e.g. sad may include only small movements. It is important to distinguish these from a neutral expression. For that reason, a set of seven expressions was recorded, containing the six basic expressions and neutral.

4.2 EMG Equipment

To measure and amplify the EMG signal, we used the Varioport III system from Becker Meditec [Med03]. This system consists of electrodes, an amplifier, a recorder and a lightwave interface (LWI). For the facial measurements reusable Ag/Ag-Cl gelled surface electrodes were employed with a diameter of 5 mm. As a ground electrode (GND) we used a disposable adhesive Ag/Ag-Cl electrode. The reusable electrodes were filled with electrolytic gel to reduce impedance between skin and electrodes.

The amplifier provides eight channels for bipolar or unipolar measurements and one ground channel (GND). The signal is digitized with a resolution of 16 bit, which corresponds to 0.033μV per unit (see technical specification in Table 4.1).

EMG Channels	8
Frequency range per channel	Channels 1-7: 0.9 - 295 Hz Channel 8: 19 - 296 Hz
Amplification factor	±1170
Range	1170 μV
A/D conversion	16 Bit
Resolution	0.033μV per Bit

Table 4.1: Specifications of VarioPort III

The recorder provides two options, directly streaming the detected EMG signal via a serial interface or storing data on an included memory card. When using the direct streaming, the maximal sampling rate is determined by the amount of used channels and transmission rate of the serial port. When all 8 channels are used the maximal sampling rate is about 600 Hz and the high-pass cutoff frequency is at about 295 Hz.
The LWI is an interface between the EMG recorder and a serial port at a PC. It internally transmits electrical signals via light impulses. This ensures that at no time the human subject is connected to a power point and it helps to minimize interferences through the power line.

4.3 Electrode Placement

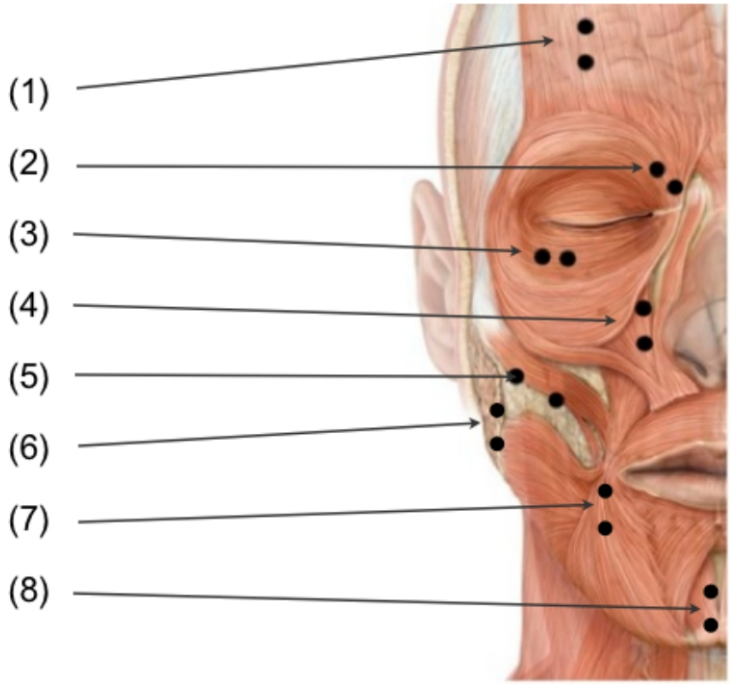

Figure 4.1: Electrode placement with channel numbers: M. Frontalis (1), M. Corrungator (2), M. Orbicularis Oris (3), M. Levator Labii (4), M. Zygomaticus (5), Masseter (6), M. Depressor Anguli Oris (7), Mentalis (8) (Background Image, [SSSbKW06])

The Varioport III system allows a maximal number of 8 bipolar electrode positions and one ground electrode. Therefore, a set of electrode positions was chosen according to [FC86] (see section 2.3). We included positions close to muscles, which are most likely to produce the recorded expressions. Figure 4.1 illustrates the electrode placement of the eight electrode pairs. The numbers represent the placement order of the selected positions, descending from the top. From now on, we will also refer to the electrode positions as channels. The exact order of each electrode pair attachment is important. When the pair is switched, the signal polarization is inverted. The GND electrode was placed on the underside of the left forearm. This location is ideal, as it contains no big muscles and is far away from the measuring site.

4.3.1 Estimated Activation Pattern

To reduce the amount of placed electrodes prior to the data recording, we analyze the estimated muscle activation on each position for the chosen facial expressions. Gosselin et al. [GKF95] analyzed the probabilistic facial activity for each of the basic expressions. They listed the mean and variance of facial motion units during each expressions. The motion units were coded in FACS as action units (AUs), which are unique observable rapid facial changes (see section 2.1.1). The definition of some AUs in FACS are based on small muscles or muscles of deeper layers [ER05].

As described in chapter 2, only a part of the activation potentials that occurs during a muscle contraction is detected with sEMG. This part originates from muscle fibers that are close enough to the measuring electrode and produce potentials strong enough to travel through the skin (see chapter 2.4). Therefore, facial sEMG is more likely to detect activations of facial regions than the activity of a single underlying muscle. Hence, we mapped the AUs described in FACS to the closest positions included in our placement. The mapping was done according to the facial muscle anatomy (Figure 2.2). Table 4.2 shows the mapping of the eight channels to the listed AUs in Table 4.3.

Channel	Electrode Position	FACS Action Unit
1	M. Frontalis	1, 2
2	M. Corrungator Supercilii	4
3	M. Orbicularis Oculi	6, 7
4	M. Levator Labii Superioris	9, 10
5	M. Zygomaticus Major	11, 12
6	Masseter	26
7	M. Depressor Anguli Oris	8, 15, 16, 25
8	Mentalis	17

Table 4.2: Mapping of FACS Action Units onto the Electrode Placement.

The estimated activation of each channel resulting from the mapping applied on the results of [GKF95] is visualized in Figure 4.2. The first graph shows the mean activation and the second graph presents the standard deviation of each channel. Figure 4.3 shows the mean activation intensity ordered by expressions. All values are scaled relative to the maximal occurring values. Musculus zygomaticus major (channel 5) has a remarkable higher activation during ”happiness” than during any

AU No.	FACS Name	Muscular Basis
1	Inner Brow Raiser	Frontalis, Pars Medialis
2	Outer Brow Raiser	Frontalis, Pars Lateralis
4	Brow Lowerer	Depressor Glabellae; Depressor Supercilli; Corrugator
5	Upper Lid Raiser	Levator Palebrae Superioris
6	Cheek Raiser	Orbicularis Oculi, Pars Orbitalis
7	Lid Tightener	Orbicularis Oculi, Pars Palebralis
8	Lips Toward Each Other	Orbicularis Oris
9	Nose Wrinkler	Levator Labii Superioris, Alaeque Nasi
10	Upper Lip Raiser	Levator Labii Superioris, Caput Infraorbitalis
11	Nasolabial Furrow	Zygomatic Minor Deepener
12	Lip Corner Puller	Zygomatic Major
13	Cheek Puffer	Caninus
14	Dimpler	Buccinnator
15	Lip Corner Depressor	Triangularis
16	Lower Lip Depressor	Depressor Labii
17	Chin Raiser	Mentalis
18	Lip Puckerer	Incisivii Labii Superioris; Incisivii Labii Inferioris
20	Lip Stretcher	Risorius
22	Lip Funneler	Orbicularis Oris
23	Lip Tightner	Orbicularis Oris
24	Lip Pressor	Orbicularis Oris
25	Lips Part	Depressor Labii, or Relaxation of Mentalis or Orbicularis Oris
26	Jaw Drop	Masseter; Temporal and Internal Pterygoid Relaxed
27	Mouth Stretch	Ptergoids; Digastric
28	Lip Suck	Orbicularis Oris
38	Nostril Dilator	Nasalis, Pars Alaris
39	Nostril Compressor	Nasalis, Pars Transversa and Depressor Septi Nasi
41	Lid Droop	Relaxation of Levator Palpebrae Superioris
42	Slit	Orbicularis Oculi
43	Eyes Closed	Relaxation of Levator Palpebrae Superioris
44	Squint	Orbicularis Oculi, Pars Palpebralis
45	Blink	Relaxation of Levator Palpebrae and Contraction o Orbicularis oculi, Pars Palpebralis
46	Wink	Orbicularis Oculi

Table 4.3: Definition of Action Units (AUs) in the Facial Action Coding System [ER05]

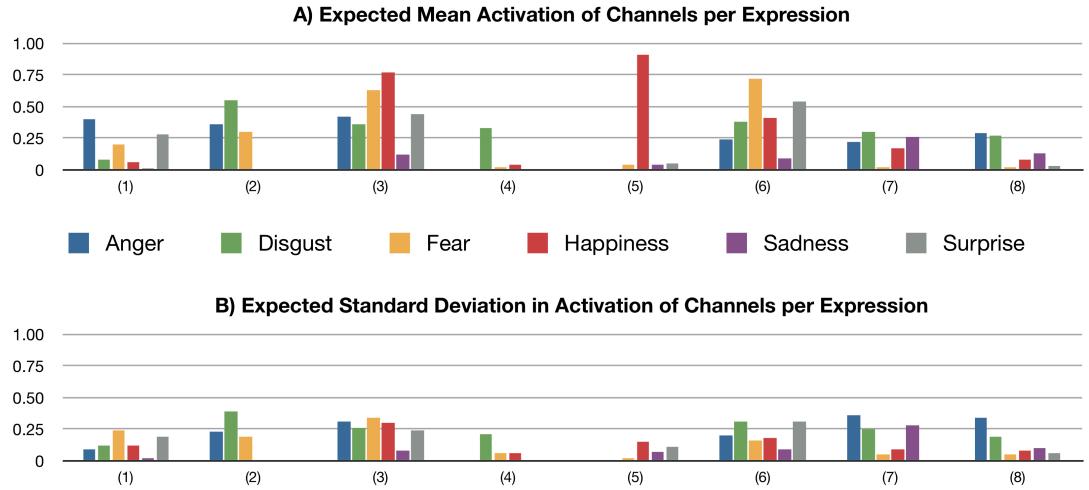

Figure 4.2: Probabilistic channel activation of mapped data from [GKF95]. a) Relative Mean b) Relative Variance. (1) M. Frontalis (2) M. Corrungator (3) M. Orbicularis Oris (4) M. Levator Labii (5) M. Zygomaticus (6) Masseter (7) M. Depressor Anguli Oris (8) Mentalis

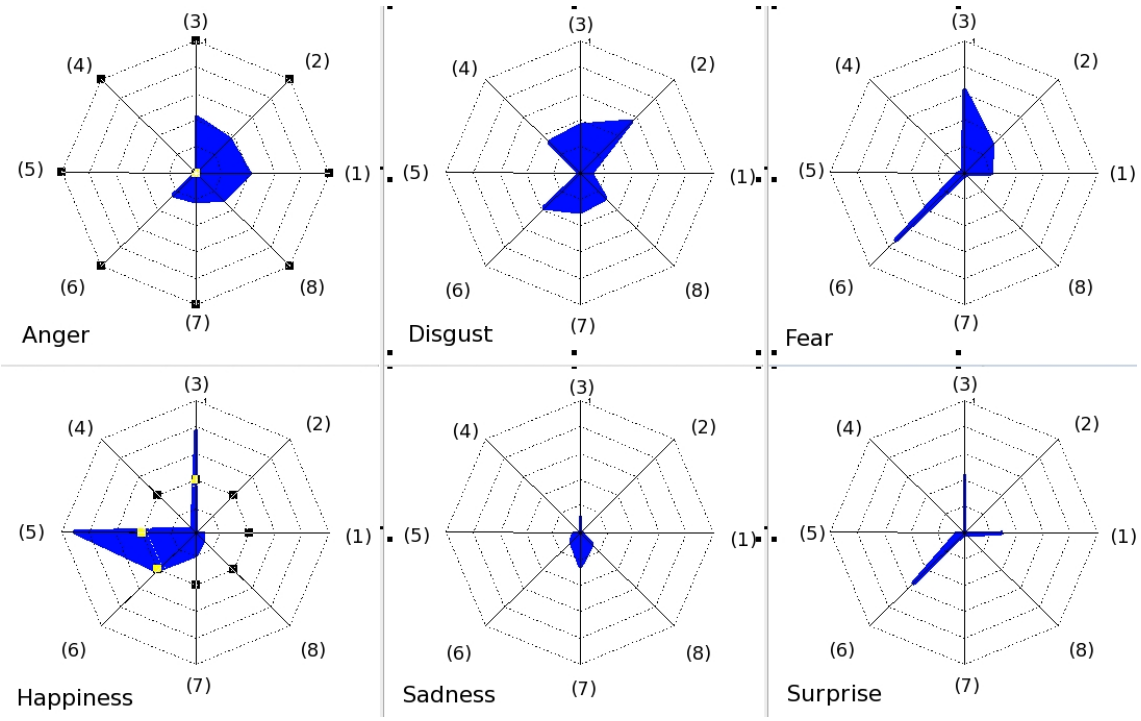

Figure 4.3: Probabilistic channel activation of mapped data from [GKF95] for each of the eight channel ordered by expression.

other expression. Channel 7 and 8 seem to have similar activation patterns for the different facial expressions, but differ in happiness and sadness. The m. corrungator (channel 2) has higher activities for anger, disgust and fear than happiness, sadness and surprise. The neutral expression is not listed in this figure, as we expect only low signals on all positions.

According to these results no clear distinguishable channel subset exists, which provides all information necessary to best classify all expressions. Therefore, to reduce electrode positions we will compute a feasible subset by a systematic reduction method (see chapter 6).

4.4 Recording Setup

This section describes the general experimental setup. Furthermore, the stimuli presentation and software used for the experiment is described. Finally, we summarize the recorded sessions.

4.4.1 General Setup

For each participant, the sEMG signals of eight electrode positions are recorded during the experiment. Similar to Chin et al. [CAG08] the participants pose different facial expressions, which are sequently indicated by visual stimuli representing one of the facial expressions.

Before starting with a detailed instruction of the participant, the electrodes are applied. The skin at the electrode positions is prepared with alcohol to reduce skin impedance and increase electrode adhesion. Then, the electrodes are filled with electrolytic gel and attached with a glue ring. The connection cables between amplifier and electrodes are affixed behind the ear to reduce irritations of the subject and to minimize potential distortions through equipment movements. The electrodes are placed in the order shown in Figure 4.1. Finally, before recording, a short introduction and an example presentation of all stimuli is given (see section 4.4.2).

During recording, the subject sits alone in a quiet room, in order to reduce environmental disturbances through distraction by other persons. The room includes a minimum of electrical equipment to reduce ambient electromagnetic noise. Each recording session consists of five blocks. Every block contains the same balanced number of samples for each of the seven expressions. The expression order is randomized to avoid training effects. Each of the five blocks is randomized in a different order.
Between blocks subjects are given a variable amount of time to relax. This is important to avoid strong differences in the subject performance in blocks due to tiredness, which may result from the monotony of the task. The length of each break is individually timed. Furthermore, a video is taken of each session.

4.4.2 Subject Introduction

In the beginning of each recording session the subject is shortly informed that facial expression will be recorded and therefore electrodes have to be placed on their face. It is explained that they have the opportunity to cancel the experiment at any time. This is again mentioned during the breaks. Then, the subject is told that at no

time any health risk will occur due to the attached electrodes. This information is necessary to minimize potential suspiciousness regarding the attached hardware.

After placing the electrodes, we present examples of all visual stimuli (see section 4.4.3) that will be displayed during the session. The subject is queried if every displayed expression is recognized and can be produced on purpose by the subject. The posed expressions do not have to match the presented picture in detail, e.g. the eyes do not have to be closed in a sad expression. In this way, the subject is posing an expression in a more natural way. During the experiment, sounds are indicating the start and stop of a expression presentation phase. The subject is asked to pose an expression until the stop-tone is played.

4.4.3 Stimuli Presentation

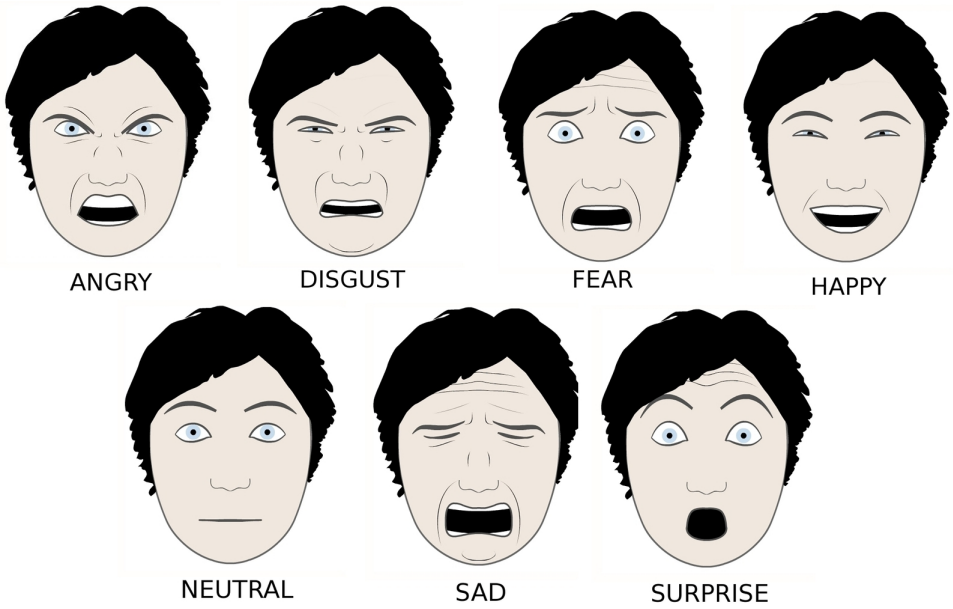

Figure 4.4: Stimuli for the seven expressions: Anger, Disgust, Fear, Happiness, Neutral, Sadness and Surprise (Created with [SF09])

.

During the whole recording session each subject is sitting on a chair in front of a TFT screen. The distance between the subject's eyes and screen is approximately one meter. The screen size measures a width of 28 cm and a height of 18 cm. To evoke the posing of the different expressions, a face demonstrating the expression and its name is displayed (see Figure 4.4). Each expression sample is presented in a cycle consisting of three phases (see Figure 4.5). In the first phase a black cross is shown in the middle of the screen for 1 second. This phase helps the subject to focus and prepare for the next expression. The focusing is also done to avoid head motions during the expression presentation. Next, a tone is played and an expression stimuli is displayed in the middle of the screen. This presentation phase lasts 4 seconds. After the presentation, a second tone is played and a grey bar is displayed for 3 seconds for relaxation. The EMG signal is recorded separately for each complete cycle in combination with an additional synchronisation channel (sync Channel), which include '1' for each frame during the expression phase and otherwise '0' (see Figure 4.5).

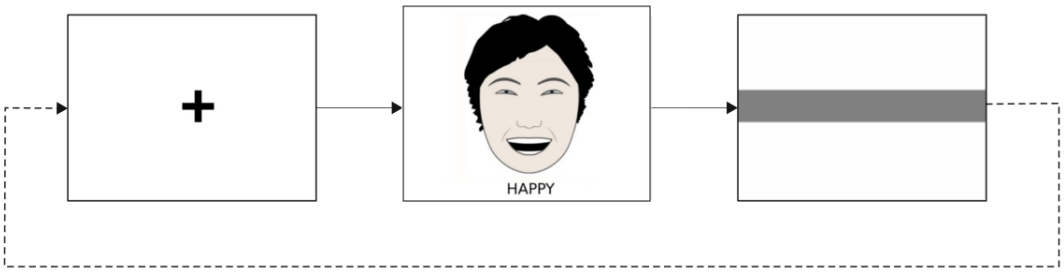

Image	Cross(1s)	Expression (4s)	Grey Bar (3s)
Sync Channel	0	1	0

Figure 4.5: Presentation Cycle: Focusing, Stimulus Presentation and Relaxation. The Sync Channel marks if a picture is displayed.

4.4.4 Software

To record the sEMG signal and to visualize the stimuli we use UKA EMG/EEG Studio v.2 [May05], which was modified by [Sch08] in a similar experimental design. This modified version was adapted to be used with the EMG equipment and to play the acoustic signal in presentations phases. It provides a visual user interface to describe session specifications, as subject and session number, the amount of used channels, sampling rate and a set of pictures.

4.5 Data Corpus

We recorded a total of three male subjects aged between 25 and 28 years. They were non-paid volunteers. All subjects self-reported to be healthy, wide-awake and right handed. Neither had taken part in a similar experiment before. None had been using medications that could affect the recordings.

A total of four sessions were recorded, with one of the three subjects being recorded twice on the same day. Between these two sessions electrodes were removed and reattached. In each session five blocks were recorded, which contained 10 samples for each of the seven expressions. In this way, a total of 350 expression samples were recorded in every session. Each expression cycle lasted 8 seconds and a block including 70 cycles about 9 minutes. The total experiment including introduction, electrode placement and breaks took about 80 minutes per session.

The EMG signal was decoded with a sampling rate of 600Hz (600 frames per second). Each expression was stored into its own subfolder, ordered by expression, block, session and stimuli number. The data was concatenated for all eight electrode channels written into a binary ".adc"-file, which is a analog to digital coded file containing the recorded EMG signal as integers. Furthermore, a synchronization file was stored that contains the output of the synchronization channel. In addition to the EMG recording a video was captured in an frontal angle with a sampling rate of 25Hz.

5. Methods

In this chapter we describe the methods to process the raw digitized EMG signal and classify different facial expressions by employing a Support Vector Machine (SVM). Several processing steps have to be performed prior to classifying a facial expression from the EMG signal.

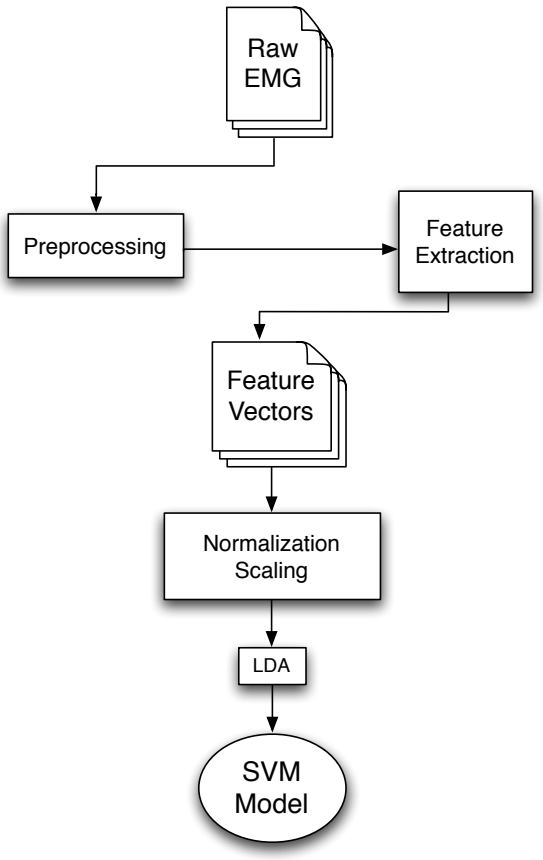

Figure 5.1: General processing chain from raw sEMG to SVM model training.

Figure 5.1 shows this general processing chain. In the first step, the signal offset is estimated and subtracted. From the offset-subtracted signal, features are extracted that describe characteristics of the signal waveform (see section 5.2). Those features are then combined to feature vectors. The features are normalized and scaled as described in section 5.3. Then, the dimensions of the feature vectors are reduced by applying a linear discriminant analysis (LDA). Finally, these reduced vectors are used to train and test an SVM model (see section 5.6).

5.1 Offset Substraction

An EMG signal has an average baseline of 0 mV when not containing any motion artifacts. This is due to the fact that the signal is the differential potential measured between electrodes. Muscle activation potentials travel bidirectionally from the motor endplate along the muscle fibers. The electrodes measure the same strength of potential difference at different times, but with reverse polarity. Therefore, the average signal has a baseline of 0 mV.

Due to the equipment, the recorded signal has a positive or negative offset. This results in a mean value that is positively or negatively shifted from zero. To remove this offset we have to determine the mean value for each channel. Due to the experiment design, each recorded sample contains a delay phase before subjects pose experiments. This delay consists of the time needed by a participant to recognize and react on a given stimulus. The reaction times in facial expressions to a stimuli range from 500 ms to 1000 ms, see Dimberg et al. [DTE00] and Moody et al.[MMMW07]. Our data confirmed this reaction time with few variations up to 1200 ms.The mean value of this 500 ms segment before the facial posing reaction can be used to approximate the channel offset. To minimize the influence of artifacts in this segment, we estimate the offset for each channel as the average value of this segment over all samples.

5.2 Feature Extraction

To recognize facial expressions, important characteristics of the sEMG signal have to be extracted. For each channel the electromyogram is a one-dimensional signal varying over time. To describe specific characteristics of this waveform, we split the signal into overlapping time segments and extract a number of features from each segment. All features are extracted in the time-domain and therefore no specific windowing function has to be used. The segments are defined by a window size and shift length. The overlap with the next segment is the difference between window size and shift length. In this section, different combinations of time domain features called *feature sets* are presented.

5.2.1 MimikFeat

The first type of feature sets is motivated by features commonly used in the prothesis control with sEMG, e.g. hand prothesis [Mik08]. This set shows a robust behavior in recognizing deliberately presented control commands [EH03].

As described earlier, the major information in facial sEMG signals occurs between 20 Hz and 250 Hz. Below 20Hz mainly motion artifacts are measured, which occur through skin movement (see chapter 2.5 for more details). To separate these

artifacts from the EMG signal, a Butterworth 20 Hz high-pass filter of order 6 is applied. The Butterworth filter is also known as *maximally flat magnitude filter*, which is maximally flat in the passband and rolls of towards zero in the stopband. No additional low-pass filter has to be applied, as the data corpus includes only frequencies up to 296 Hz. Additionally, a notch filter could be applied in order to suppress specific frequency ranges, e.g. to reduce power line noise at 50 Hz. But notch filters do not filter selectively at a sharp frequency range but also influence neighboring frequencies. Futhermore, the major amount of EMG signal activity is located in a lower frequency range, in which the power noise is located. In this way, the notch filtering would remove important information. Hence, we do not apply any notch filtering following the recommendation of [RHY06]. The filtered signal for each segment with the length of N_{sig} frames and the filter order of b has the form:

$$x^{filt} = \{x_{filt}(b), x_{filt}(b+1), .., x_{filt}(N_{sig}-1), x_{filt}(N_{sig}) \tag{5.1}$$

On this filtered signal we define our first feature set *MimikFeat1*. The features we extract from each segment are the root mean square (RMS) $\bar{X}_{rms}$, gradient of the RMS $\Delta\bar{X}$, the waveform length l, zero crossing rate z and slope sign change s.

The root mean square(RMS) describes the average activation level of the segment:

$$\bar{X}_{rms} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \tag{5.2}$$

The gradient of RMS signal is a measure of the activity development between segments:

$$\Delta\bar{X} = \bar{X}_{rms}(i) - \bar{X}_{rms}(i-1) \tag{5.3}$$

The waveform length of the filtered signal is a measure of the signal complexity in each segment. The resulting value combines waveform amplitude, frequency and duration within a single parameter [EH03]:

$$l = \sum_{i=2}^{N} |\Delta p(i)| \text{ with } \Delta p(i) = p(i) - p(i-1) \tag{5.4}$$

The zero crossing rate on the filtered signal is a measure of the signal frequency:

$$z = \sum_{i=2}^{N} z(i) \text{ with } z(i) = \begin{cases} 1 & \text{for } x_{filt}(i) \cdot x_{filt}(i-1) < 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.5}$$

The slope sign change on the filtered signal is another measure of frequency:

$$s = \sum_{i=2}^{N-1} p[i] \text{ with } p[i] = \begin{cases} 1 & \text{for } \Delta p_i \cdot \Delta p_{i+1} < 0 \\ 0 & \text{otherwise} \end{cases} \text{ with } \Delta p_i = p(i) - p(i-1) \tag{5.6}$$

The resulting MimikFeat1 feature vector for each segment has the from:

$$\text{MimikFeat1} = [\bar{X}, \Delta\bar{X}, l, z, s] \tag{5.7}$$

In addition to the feature set of MimikFeat1 we define a second feature set *MimikFeat2*, extracting additional features on low frequencies below 20 Hz. Including information about motion artifacts in the classification task may lead to better classification. To create the low frequency signal a Butterworth low-pass filter of 20 Hz was applied to the raw EMG signal.

From this low frequency signal we extract a measurement for signal activation and inter-segment development by calculating the RMS $\bar{X}_{low}$ and ΔRMS $\Delta\bar{X}_{low}$.

The resulting feature vector is stacked on MimikFeat1 vector and has the form:

$$\text{MimikFeat2} = [\bar{X}_{low}, \Delta\bar{X}_{low}, \bar{X}, \Delta\bar{X}, l, z, s] \tag{5.8}$$

5.2.2 TD

The second type of feature sets is called *TD* (time domain) and showed good performances in speech recognition based on sEMG (Wand et al. [WS09]). First, the offset-subtracted signal is split into two parts that approximately contain lower and higher frequencies. This is realized by calculating an averaged signal and its difference to the original signal. A two times 9-Point averaging filter is applied to smooth the signal:

$$w[n] = \frac{1}{9}\sum_{n=-4}^{4} v[n], \text{ where } v[n] = \frac{1}{9}\sum_{n=-4}^{4} x[n] \tag{5.9}$$

.

The averaging is comparable to a low frequency filter because mainly higher frequent changes are suppressed in the signal. The resulting waveform is referred to as "low frequency" (LF). The LF mainly describes distortions of the sEMG signal like shifting of the mean due to motion artifacts. Including these values can provide information about the direction and the energy of facial motions.

The second signal part, the "high frequency" (HF), is defined as the difference between the raw signal and the LF:

$$p[n] = x[n] - w[n] \tag{5.10}$$

On HF the rectified values are calculated (Equation 5.11):

$$r[n] = \begin{cases} p[n] & \text{for } n \geq 0 \\ -p[n] & \text{for n} < 0 \end{cases} \tag{5.11}$$

For each time segment we extract the following features from the two signal parts:

Mean of LF:

$$\bar{w} = \frac{1}{N}\sum_{i=1}^{N} w[i] \tag{5.12}$$

The mean of rectified HF:

$$\bar{r} = \frac{1}{N}\sum_{i=1}^{N} r[i] \tag{5.13}$$

The power of LF

$$P_w = \sum_{i=1}^{N} |w[i]|^2 \tag{5.14}$$

The power of HF

$$P_r = \sum_{i=1}^{N} |p[i]|^2 \tag{5.15}$$

The zero crossing rate of HF

$$z = \sum_{i=2}^{N} z(i) \text{ with } z(i) = \begin{cases} 1 & \text{for } p[i] \cdot p[i-1] < 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.16}$$

To include context information about the current segment a context filter is introduced, joining neighboring feature vectors. This is realized by stacking the c previous and c successive vectors onto the current vector:

$$f_{context} = [f_{-c}, f_{-c+1}, .., f_0, .., f_{c-1}, f_c] \tag{5.17}$$

Combining these features and the context filter we define the TD feature set *TD0*. Hereby, "TD" stands for time-domain feature and a context filter of 0. The resulting feature vector for each segment has the form:

$$\text{TD0} = [\bar{w}, \bar{r}, P_w, P_r, z] \tag{5.18}$$

TD0 is analog to the feature set "f2" in the study of Jou et al. [JSW+06]. In this study on speech recognition using facial surface EMG, they compared several feature combinations. The best performing feature set in this study included a context filter of width 5. This corresponds in our study to *TD5* and has the form:

$$\text{TD5} = [TD0(i-5), TD0(i-4).., TD0(i+4), TD0(i+5)] \tag{5.19}$$

In our study several combinations of TD features will be used according to this definition.

5.3 Feature Normalization and Scaling

The range and distribution of a specific feature is not necessarily an indication for its relevance in classification. Hence, to promote better classification performance, we need to normalize and scale the feature vector data. The parameter for normalization and scaling are estimated on the training data.

5.3.1 Normalization

In order to normalize our features we use *z-score transformation* commonly employed in statistics. The global mean $\bar{x}_i$ and standard deviation σ_i is estimated on each feature i over all training data. Through this normalization step the resulting standard deviation ranges from -1 to 1 and a mean value for each feature of 0.

$$x_i^{norm} = \frac{x_i - \bar{x}_i}{\sigma_i} \tag{5.20}$$

5.3.2 Scaling

On the normalized features a linear scaling is applied. We extract the maximum and minimum values in the training data for each feature. These values are defined as the maximum max_i^{norm} and minimum min_i^{norm} value limits for each feature i. Only values in this range will be accepted in future data. Potentially higher or lower future values will be set to these extreme limits by:

$$x_i^{norm'} = \begin{cases} max_i^{norm}) & \text{for } x_i^{norm} > max_i^{norm} \\ x_i^{norm} & \text{for } x_i^{norm} \in [min_i^{norm}, max_i^{norm}] \\ min_i^{norm} & \text{for } x_i^{norm} < min_i^{norm} \end{cases} \tag{5.21}$$

The values of each feature are scaled by division by its maximum absolute value of the estimated borders. This results in values in the range between -1 and 1.

$$x_i^{scaled} = \frac{x_i^{norm'}}{max(|min_i^{norm}|, |max_i^{norm}|)} \tag{5.22}$$

5.4 Feature Reduction

The dimensionality of feature vectors can be very high, depending on the number of features and channels. Therefore, a dimension reduction method is applied before training and classification. This is realized by emphasizing those features and channels that increase the separability of classes. We employ a Linear Discriminant analysis (LDA), which is a popular method for feature reduction in classification problems with a large number of features.

The LDA estimates a linear transformation from the original feature space into a lower-dimensional space and maximizes the separability between data points of different classes. The idea behind LDA is to reduce the feature dimensionality with a minimal loss of discrimination information. The mapping is calculated by minimizing the inner-class variance and maximizing the distance between the mean

of classes. In order to employ an LDA, each datapoint of the used dataset has to be labeled with its class affiliation.

Consider a discriminative problem with C classes defined by their mean vectors μ_i with $i = 1, 2, .., C$. The feature vectors to be classified have the form $x = (x_1, x_2, ..., x_N)^T$ with N being the number of features. Based on this we can define the within-class scatter S_w and the between-class scatter S_b where μ is the mean over all feature vectors:

$$S_b = \sum_{i=1}^{C} (\mu_i - \mu)^T \cdot (\mu_i - \mu) \tag{5.23}$$

$$S_w = \sum_{i=1}^{C} \sum_{x \in C_i} (x - \mu_i)^T \cdot (x - \mu_i) \tag{5.24}$$

The transformation matrix T for the mapping is computed by minimizing the within-class scatter and maximizing the between-class scatter. This is done by maximizing the criterion γ:

$$\gamma = \frac{M^T S_b M}{M^T S_w M} \tag{5.25}$$

This maximization can be estimated by solving the generalized eigenproblem:

$$S_b v = \lambda S_w v \tag{5.26}$$

Thereby, the columns of the transformation matrix M are the eigenvectors v. In this study, the LDA implementation of the dimension reduction toolbox by van Maaten [vdMH08] was used.

5.5 Electrode Reduction

To reduce the number of electrode positions, we use a Sequential Backward Channel Reduction (SBCR), which is an analog algorithm to Sequential Backward Feature Selection (SBFS). The algorithm follows a greedy approach. Beginning with a full channel set a criterion function is calculated for each possible subset, which can be build by removing one channel. The subset with the highest criterion result is kept. Analog to this, in each reduction step one channel is removed. This is continued until a maximum criterion is reached or only one channel remains.

5.6 Support Vector Machines

To train and classify patterns in the EMG signal, we employ Support Vector Machines (SVMs). The following sections give a short introduction into the basic principle of SVMs and the application on non linearly separable data. For a more detailed introduction refer to Burges [Bur98] and Boser [BGV92].

SVMs estimate a hyperplane that optimally separates a given data set. An advantage of SVM is in using kernel functions to map data into a higher (up to infinite) dimensional space, in which a linear discrimination can be trained. This follows the principle that with an appropriate mapping to a sufficiently high dimension, data from two classes can always be separated linearly by a hyperplane [DHS01]. Therefore, SVMs can be applied to solve complex non-linear classification problems.

5.6.1 Basic Principle

Given data from two classes and a mapping Φ into an appropriate space, in that the data becomes linearly separable. In this space SVMs estimate the hyperplane $H0$ separating the classes with the maximum margin (see Fig.5.2). This hyperplane is defined by a normal vector w and its distance b to the coordinate origin:

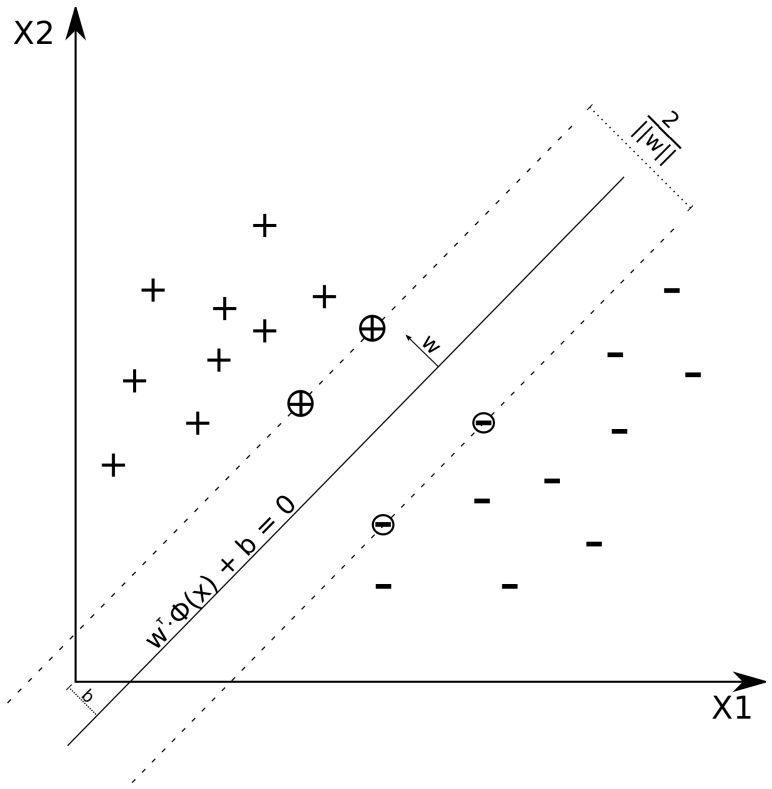

Figure 5.2: Example of an SVM hyperplane with maximum margin trained on data samples of two classes.

$$w^T \cdot \Phi(x) + b = 0 \tag{5.27}$$

Given linearly separable data, two hyperplanes $H1$ and $H2$ exist parallel to $H0$(5.27), where each crosses through the closest data points of one of the two classes. All points within the margin are defined as support vectors (SV). SVs provide the highest information about the differentiability of a data set as they define the smallest differences between classes.

SVMs estimate the separating hyperplane $H0$ with the maximum distance to $H1$ and $H2$ by minimizing the term $\frac{1}{2}w^T w$ and solving the quadratic optimization problem

5.28. Here, the linearly separable data is defined as $TR = (x, y)$ with $x \in \mathbb{F}$ and $y = \{1, -1\}$, $\mathbb{F}$ is the feature space and y the data labels. SVMs calculate a unique solution for the same set of SVs regardless of the additional data points.

$$\begin{aligned} &\text{Minimize } \frac{1}{2} w^T w \\ &\text{subject to } y_i(w^T \cdot \Phi(x_i) + b) \geq 1, \forall (x_i, y_i) \in TR \end{aligned} \tag{5.28}$$

5.6.2 Soft Margin

To avoid generalization loss through outliers in the data or measuring errors, a soft margin is introduced, allowing a restricted number of misclassifications within the margin. The optimization problem 5.28 then has the form:

$$\begin{aligned} &\text{Minimize } \frac{1}{2} w^T w + C \cdot \sum_{i=1}^{N} \xi_i, \\ &\text{subject to } y_i(w^T \cdot \Phi(x_i) + b) \geq 1 - \xi_i, \\ &\forall (x_i, y_i) \in TR, \xi_i \geq 0 \end{aligned} \tag{5.29}$$

Where ξ_i are slack variables measuring the degree of misclassification and the parameter C is a penalty parameter through which the amount of allowed misclassification errors is controlled.

To overcome the problem of high computational effort through (5.29), the term is transformed into a dual problem, which means that the optimization term only depends on the scalar products of $\Phi(x_i)$. Then, using Lagrange Multipliers, the weight vector w can be written as a linear combination of training samples [Bur98]:

$$\begin{aligned} &w = \sum_{i=0}^{N} \alpha_i y_i \Phi(x_i) \\ &\text{where } \alpha_i = 0 \text{ for all } x_i \text{ with } y_i(w^T \cdot \Phi(x_i) + b) > 1 \end{aligned} \tag{5.30}$$

Only the SVs are left over in the equation and the scalar product. This allows the application of kernel functions. This is also commonly know as the ”kernel trick”, which allows the usage of SVMs for non linear problems while not heavily increasing its computational effort.

5.6.3 Multi-Class Classification

SVMs were developed to solve binary classification problems. Therefore, a modification has to be defined to apply them to multi-class problems. In this work a one-versus-one approach was used. In this approach, an SVM is trained for each possible pair of classes. This results in a total number of $\frac{k(k-1)}{2}$ SVMs to be trained, where k is the number of classes.

A data point is classified by majority decision. In case of two classes having identical votes, the one with the smallest class index is selected. For this work we used the LibSVM 2.89 library, for detailed implementation description refer to Chang et al. [CL09].

5.6.4 Radial Basis Function

Different kernel functions can be used to map data into an appropriate space. The most common kernels are linear, sigmoid, polynomial kernels and radial basis functions (RBF). In this work we use RBFs. Compared to linear kernels this functions can transform into higher dimensional spaces and are applicable on non linear separable data. Sigmoid kernel functions tend not to achieve better results than RBFs and may not be positive definite [LL03] and for polynomial kernels more parameters need to be estimated. The RBF kernel has the form:

$$K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j) = e^{-\gamma ||x_i - x_j||^2}, \ \gamma > 0 \tag{5.31}$$

The value of the parameter γ determines the width of the RBF.

5.6.5 Parameter Selection

The parameters C and γ are not estimated during the training process and have to be set beforehand. Following the approach of Hsu et al.[CHCL03], a grid search is used to estimate the best pair of C and γ. Thereby, the SVM model is trained on training data and the classification rate is estimated on optimization data, which is data that was not part of the data used to train the tested SVM, and combinations of parameters are tested with $C = \{e^{-5}, .., e^{15}\}$ and $\gamma = \{e^{-15}, .., e^{3}\}$. The set of C and γ is selected that results in the highest classification rate.

6. Evaluation

6.1 Expression Recognition

In this chapter we describe the experiments that we conducted on our data corpus and their results. The general structure of the developed system is presented and the experiments to determine the best feature set and parameters are introduced. Finally, we analyze the reduction of electrode positions.

6.1.1 System Structure Description

The general structure of the classification system, we conducted the experiments with, is shown in Figure 6.1. It consists of two major parts: the feature extraction with assignment in training and test sets (Figure 6.1(A)) and the SVM training and classification, including normalization and dimensional reduction (Figure 6.1(B)).
In part (A), the EMG signal is preprocessed and the samples are randomly assigned to different sets for training and testing. Each set contains a balanced number of samples for each expression. The randomization minimizes the impact of potential session inherent trends in the signal like muscle fatigue and changes in the expression posing over time. This assignment of sample data is stored and used for all conducted experiments to guarantee comparability between the experimental results. The 50 recorded samples for each expression were divided into 30 expression samples for training and 20 samples for testing. Then, feature vectors are extracted for each expression sample, dependent on the used window size and feature set.

The actual training and testing of the classification system is shown in Figure 6.1.B. On the training data, the parameter for normalization, scaling and the LDA mapping are calculated and applied on the whole data set. Based on this preprocessed data and the expression affiliation of each feature vector, an SVM model is trained. The classification is realized as the class prediction result of a feature vector by the trained model.

The system was implemented in MATLAB (Version R2009a, Math Works Inc., Natick, Massachusetts) using libSVM [CL09] and the Dimension Reduction Toolbox [vdMH08].

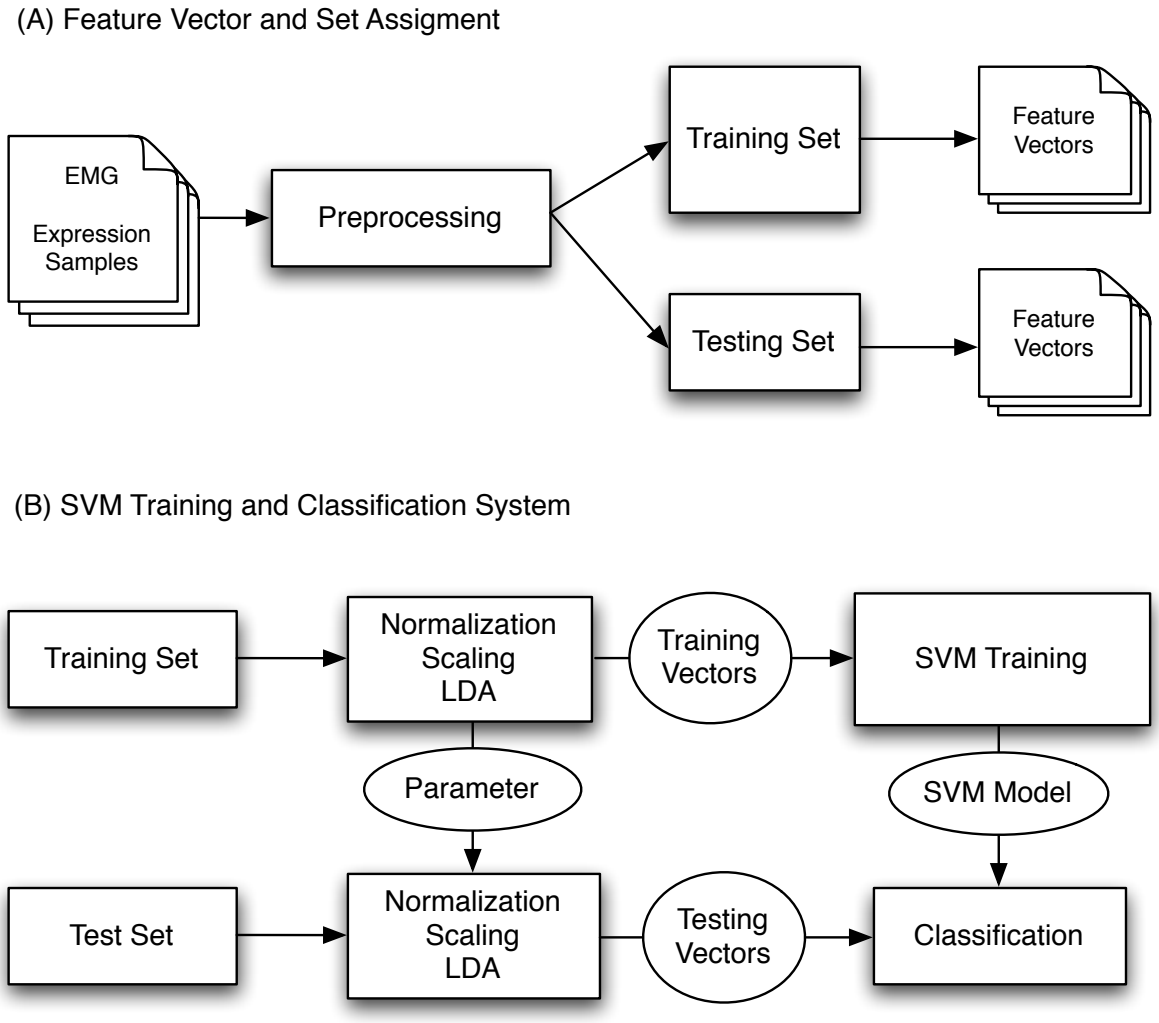

Figure 6.1: General structure diagram of the evaluation system.

6.1.2 Feature Set Selection

In this experiment we want to determine the best feature set of our sets to recognize facial expressions. In this section we analyze the general performance of TD0 in comparison with MimikFeat1 and MimikFeat2. We compare different classification performances for the different feature sets on different window sizes. For each session the SVM model was trained on the previously assigned training set and tested on the test set.

First, we analyze the expression recognition rate (ERR) based on each training example by majority vote. A test sample is counted as correctly classified if more than 50% of the feature vectors of this sample are correctly classified. Figure 6.2 shows the subject average ERR results. In order to minimize the impact of each subject, the average ERR is calculated, as the average of subject A, C and the average result of the two sessions of subject B.

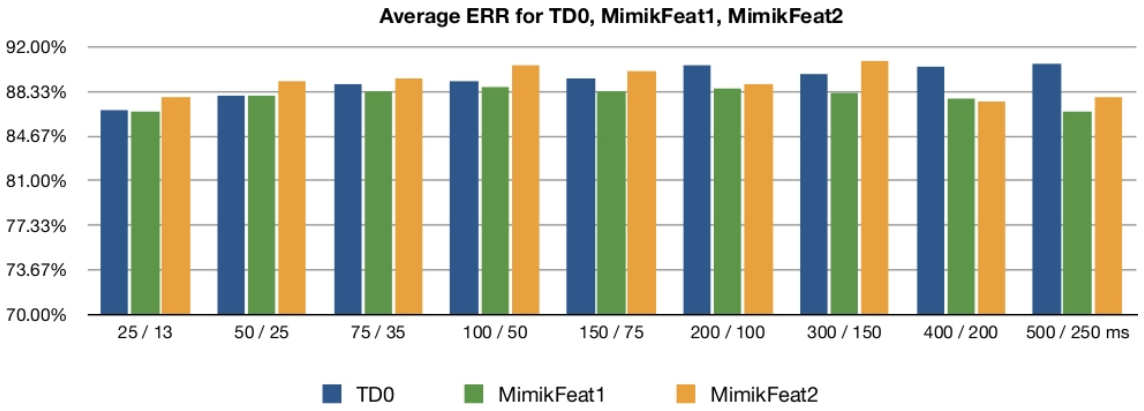

Figure 6.2: Average expression recognition rate (ERR) of TD0, MimikFeat1 and MimikFeat2 on different window sizes.

The results show an advantage of MimikFeat2 performing better in window sizes up to 150 ms and for 300 ms. For window sizes of 200, 400 and 500 ms, the ERR for TD0 is clearly better than MimikFeat1 and MimikFeat2. TD0 shows a clear trend of increasing ERR in relation to growing window sizes. For detailed results of ERR for each subject and session, refer to Table A.1 in Appendix A.

The next step is to analyze the recognition rate on the correctly classified feature vectors (VRR). VRR is the rate of correctly classified feature vectors in relation to all classified feature vectors. Figure 6.4 displays the average results for the three feature sets. There is a clear tendency of increasing VRR in relation with larger window sizes.

In the VRR results TD0 performs best and MimikFeat2 is slightly better than MimikFeat1. The average result shows a consistent behavior for window sizes of 50 to 500 ms. For segment length of 25 ms, MimikFeat2 shows a marginal advantage compared to TD0 with a difference of about 1.7%. All feature sets performed with an average VRR about 82%, which is significantly better than chance level of 14.3%. A detailed results of VRR for each subject and session, is listed in Appendix A in Table A.2.

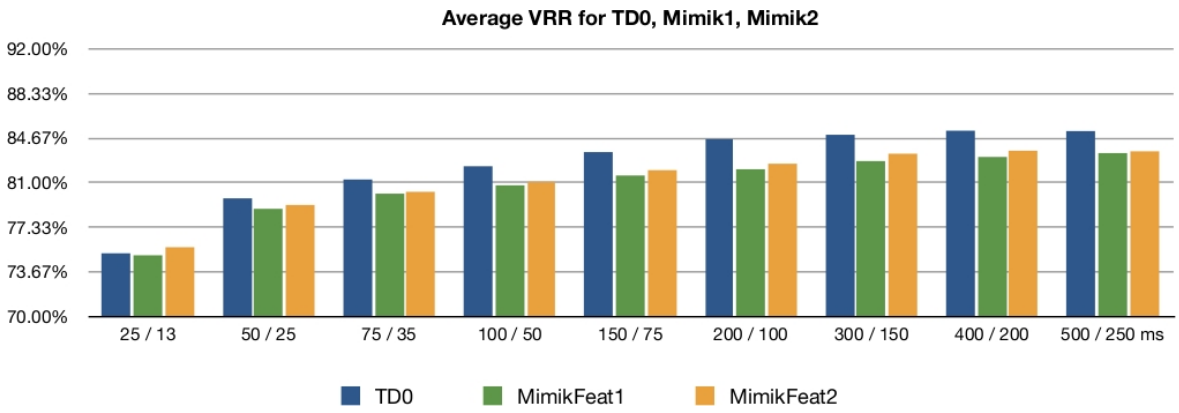

Figure 6.3: Average single feature vector recognition rate VRR of TD0, MimikFeat1 and MimikFeat2 on different window sizes.

On basis of the better performing TD0 and good results of the MimikFeat sets, we introduce combinations of the two feature set types. We add the two main different features of MimikFeat to TD: the Slope Sign Change and Waveform Length. The resulting three feature set combinations are: TD-Slope, TD-Waveform and TD-Slope-Waveform.

Testing these combined feature sets resulted in further improvements in both ERR and VRR. Figure 6.4 presents the average recognition rates of all TD0 features. TD-Slope-Waveform improves the recognition results for all tested window sizes in both ERR and VRR. The VRR of TD0-Slope-Waveform compared to TD0 is improved by 1.2 - 1.75%, which is an relative enhancement between 4.9% to 12.6%. For all window sizes, the TD variants outperformed the MimikFeat sets. Therefore, in further analysis we will determine the best set of TD variants.

6.1.3 Window Size and Context Length

In this experiment, we analyze the influence of window size on the recognition rate. Following the approach of Jou et al. [JSW+06], we also apply different context filters to the introduced TD variants (TD, TD-Slope, TD-Waveform, TD-Slope-Waveform).

The average subject recognition results for all TD variants and context lengths are presented in Appendix A Table A.3 and A.4. For each feature set, the tables show the ERR and VRR results, listed by window size and context filter.

The results show a clearly increasing recognition rate with increasing window size before the accuracy drops at large window sizes (Table A.3). This may be caused by a lower amount of feature vectors used for training and testing. The number of vectors extracted per expression sample is defined by the window shift and the total signal length, which is the amount of signal L used to generate a feature vector. The total amount of signal needed to create a feature vector is determined by:

$$L = WindowSize + 2 \cdot ContextLength \cdot ShiftSize \tag{6.1}$$

The best average ERR was reached depending on the context filter at window sizes between 75 ms to 400 ms. A trend can be identified that higher recognition rates are

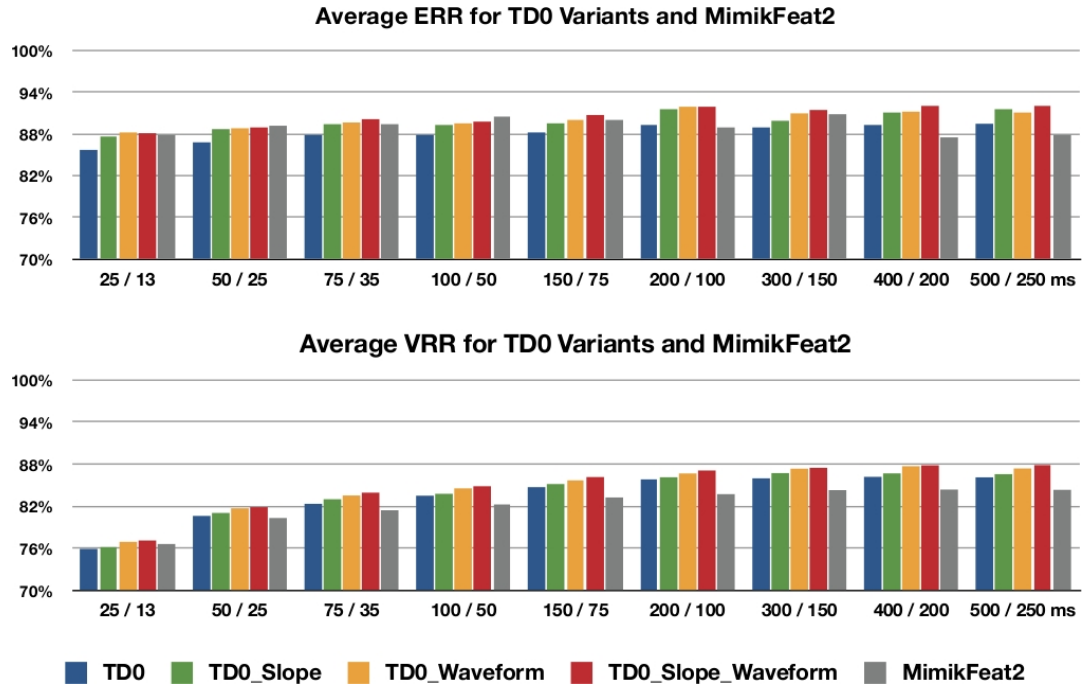

Figure 6.4: Average ERR and VRR of TD0, TD-Slope, TD-Waveform, TD-Slope-Waveform and MimikFeat2 on different window sizes.

related to larger window sizes and amount of used signal. The total time of signal length used to produce the best performing feature vectors varied between 250 to 500 ms (Table A.5).
The ERR of the feature set TD1-Waveform combined with a window size of 200 ms and TD0-Slope-Waveform combined with a window size of 400 ms, provided the best ERR. To generate the feature vectors both sets use a signal length of 400 ms (see Table A.5). The TD1-Waveform does not show a constant performance over time. The employment of context filters did not result in remarkable improvements in recognition performances compared to the time used to create the feature vectors.

6.1.4 Specific Expression Recognition Characteristics

In this section the class specific ERR is analyzed. For each expression class, we investigate the recognition results regarding to a correct classification and their confusion with a different expression. We calculate the confusion matrix C_{FS} defined by equation 6.2. Hereby, each column i represents the classification results for test data of class i.

$$C_{FS} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,k} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k,1} & a_{k,2} & \cdots & a_{k,k} \end{pmatrix} \text{ with } a_{i,j} = \frac{\text{test data of class i classified as class j}}{\text{number of test data of class i}} \tag{6.2}$$

The system details used for this task were:

- Sets(randomized samples for each expression): 20 Test Set, 30 Training Set

- Feature Set: TD0-Slope-Waveform
- Window Size: 400 ms
- Shift Size: 200 ms
- Feature Reduction through LDA to $k-1$ Dimensions
- $C = 1$ and $\gamma = \frac{1}{k}$ with $k = 7$, the number of expression classes

As determined in the previous section TD0-Slope-Waveform and a window size of 400 ms performed best. For this, Table 6.1 presents the average confusion matrix over the test sets of all sessions.

Thereby, the expression "happiness" has the highest recognition rate with an average of 98.91%. Analyzing the sEMG signal showed that this expression, in comparison to the other, has a clearly distinguishable high activation in m. zygomaticus major.

		Prediction [%]						
		Anger	**Disgust**	**Fear**	**Happ.**	**Neutral**	**Sadn.**	**Surpr.**
Target	**Anger**	**82.03**	9.06	1.56	0.78	0.31	5.00	1.25
	Disgust	7.34	**77.5**	3.59	0	0	10.63	0.94
	Fear	1.25	3.44	**83.28**	0	0.63	2.03	9.38
	Happiness	0	0.16	0	**98.91**	0	0.47	0.47
	Neutral	0	0.16	1.25	0	**92.66**	0	5.94
	Sadness	1.88	13.59	1.88	0.31	1.25	**81.09**	0
	Surprise	0	0.78	4.22	0.16	3.91	0.31	**90.63**

Table 6.1: Average confusion matrix over the test sets of all sessions, with TD0-Slope-Waveform and window size 400 ms.

The second best classified expression is "neutral", with an average of 92.66% and a small confusion distributed over all other expressions lower than 4%, while "disgust" is the most often misclassified and gets mainly confused with "anger" and "sadness". All three expressions include the lowering of the eyebrow. Furthermore, there is a confusion between "fear" and "surprise", which can be explained by the fact that both expressions joining eyebrow rising and mouth opening. In comparison to "happiness" the other expressions show higher confusion with each other, as they do not include motions of such unique facial areas.

6.1.5 SVM Parameter Optimization

In this section we will session-specifically optimize parameters of the SVM. There are two kernel parameters, which can be modified to optimize the classification results of an SVM with RBF kernel: the penalty parameter C and the kernel parameter γ. The program structure used for this optimization task is shown in Figure 6.5. The best pair of C and γ is estimated by applying a grid search (see chapter 5), as proposed by Hsu et al. [CHCL03] with:

$$C \in \{e^{-5}, e^{-3}..., e^{25}\} \text{ and } \gamma \in \{e^{-15}, e^{-13}, .., e^{-5}\} \tag{6.3}$$

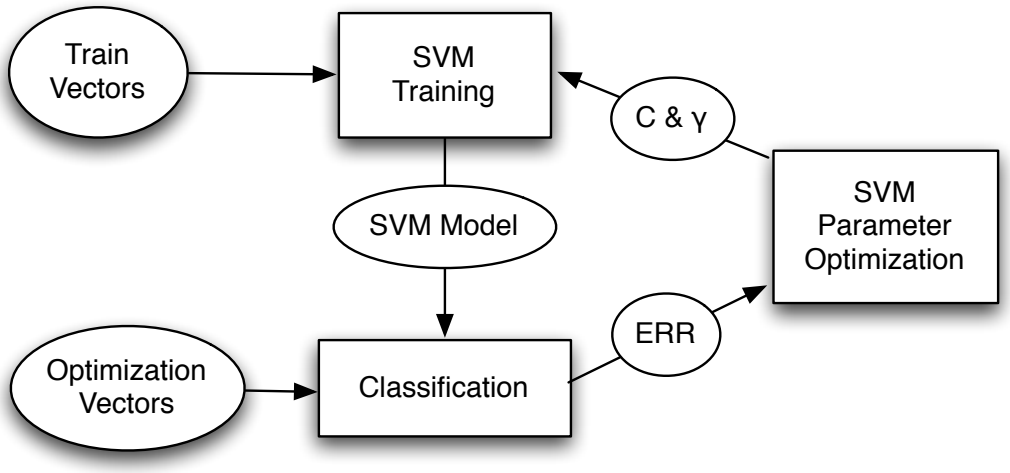

Figure 6.5: SVM parameter optimization structure.

First we divide for each expression the 30 samples of the training set into 20 samples for training and 10 samples for optimization. Thereby, all parameters for normalization and scaling and the LDA mapping are exclusively estimated on the training set. In the optimization task, the SVM model is trained on the training set and tested on the optimization set. The resulting ERR on the optimization set is used to determine the best parameter combination. Based on the best combination of C and γ the final ERR result is calculated on an SVM model, which is tested on the test set and trained on the whole training set, including all 30 examples of each expression.

The system details used for this task were:

- Sets(randomized samples): 20 Testing, 10 Optimization, 20 Training
- Feature Set: TD0-Slope-Waveform
- Window Size: 400 ms
- Shift Size: 200 ms
- Feature Reduction through LDA to $k-1$ Dimensions
  with $k = 7$, the number of expression classes

Table 6.2 shows the baseline ERR, optimized ERR and best parameters. The optimization of the RBF kernel parameters shows a slight enhancement of 0.4%, which results from a small increase of the ERR for subject A and C and a decreased ERR for the two sessions of Subject B. Overall, the parameter optimization does not provide significant enhancement to our system.

Data Set	Baseline [%]	Optimization [%]	C	γ
A	96.43	97.86	5	-10
B1	85.71	82.86	9	-5
B2	88.57	87.86	0	-3
C	91.43	92.86	2	-6
Subject Average	**91.67**	**92.02**		

Table 6.2: SVM parameter optimization results for the recorded sessions, with TD0-Slope-Waveform

6.1.6 Session and Subject Independent Classification

To determine the generalization capability of our method, we analyze the performance in subject and session independent tasks in this section.

The system details used for this task were:

- Sets(randomized samples for each expression): 20 Testing, 30 Training
- Feature Set: TD0-Slope-Waveform
- Window Size: 400 ms
- Shift Size: 200 ms
- Feature Reduction through LDA to $k-1$ Dimensions
- $C = 1$ and $\gamma = \frac{1}{k}$
  with $k = 7$, the number of expression classes

For this task, we use 30 samples for each expression for training and 20 for testing. We first train and test the system on all possible pairs of session combinations, e.g. training on training samples of A and testing on the testing samples of A, B1, B2 or C . Additionally, we analyze the performance of combined training sets of different subjects on each test set, e.g. combining training data of A,B1 and testing on testing samples of A, B1, B2 or C. Table 6.3 shows the ERR results for both tasks.

The results in the session independent tasks show low ERR, whereas the classification of data from the same subject but different sessions performs well, but still worse than in the session dependent task. Remarkable is that this data was captured with repositioned electrodes during the two sessions. In this way, the exact electrode positions slightly differ between sessions. The results show that the replacement of electrodes has a smaller impact on the classification results than individual differences between the subjects.

The combined training task show decreased classification results in cases where data of a tested session was included in the combined training. In all cases, the ERR increases on test data of unseen subjects, e.g. the combined training on A and C and testing on B2 results in an ERR of 50.71% in comparison to 20% when training on C and testing on B2.

Session Independent Task - ERR [%]					
		Test Data			
		A	**B1**	**B2**	**C**
Train Data	**A**	95.71	28.57	25.00	15.00
	B1	32.14	86.43	72.86	30.71
	B2	20.71	69.29	87.14	32.86
	C	23.57	27.14	30.00	93.57

Combined Training Task - ERR [%]					
		Test Data			
		A	**B1**	**B2**	**C**
Train Data	**A, B1**	92.86	79.29	67.14	27.86
	A, C	91.43	47.86	50.71	85.00
	B1, C	36.43	82.14	72.86	87.86
	A, B1, C	90.71	76.43	71.43	87.14

Table 6.3: ERR results for session independent training and testing.

6.1.7 Reduction of Electrode Positions

In this study we used 8 bipolar electrode positions. The placement of such a large numbers of electrodes in a human face is not feasible in applications outside laboratories or medical environments. Furthermore, the electrodes can restrict or influence a subject's behavior. Also the placement process is complex and time-consuming. Hence, the reduction of measuring electrode positions (channels) to a minimum is essential.

For this task, we applied the sequential backward channel reduction(SBCR) described in section 5.4. To obtain the best possible position reduction, we combined the results over all sessions, by using the average ERR over all sessions as the maximization criterion.

	SBCR ERR[%]				
Channels	**Subject Average**	**A**	**B1**	**B2**	**C**
1 2 3 4 5 6 7 8	**92.02**	95.71	86.43	87.14	93.57
1 2 3 5 6 7 8	**92.14**	97.86	84.29	87.14	92.86
1 2 3 5 6 7	**91.31**	97.14	83.57	85.71	92.14
1 2 5 6 7	**89.17**	96.43	77.86	84.29	90.00
1 2 5 6	**87.86**	95.00	77.14	81.43	89.29
1 2 5	**81.79**	86.43	69.29	75.71	86.43
1 5	**76.07**	81.43	66.43	71.43	77.86
5	**61.67**	63.57	52.86	55.71	67.14

Table 6.4: SBCR results on a session combined SVM model and ERRs for single session SVMs, trained and tested on the channel subset.

The system details used for this task were:

- Sets (randomized samples): 20 Testing, 30 Training
- Feature Set: TD0-Slope-Waveform
- Window Size: 400 ms
- Shift Size: 200 ms
- Feature Reduction through LDA to $k-1$ Dimensions
- $C = 1$ and $\gamma = \frac{1}{k}$ with $k = 7$, the number of expression classes

The results of SBCR in Table 6.4 show a slightly increasing recognition rate, reducing from 8 to 7 channels by removing the m. labii superioris position. This increase is mainly caused by the ERR result of subject A, by constant results for the second session of subject B (B2) and small loss of 0.15% in ERR by the first session of subject B1 and 0.71% by subject C.

A further reduction to four positions does still provide high recognition accuracies with an subject average of 87.86%. This subset includes the two positions of m. corrungator and m. frontalis on the forehead and m. zygomaticus major and masseter on the cheeks. These channels include mainly information about mouth opening, upwards pulling of the mouth corner and eyebrow movements.

This subset achieved the subject average confusion matrix shown in Table 6.5. "Anger", "fear", "happiness" and "surprise" are recognized with an ERR over 77%. "Disgust" is classified with 57.08% and mainly misclassified as "anger" with 25.42% and "sadness" with 13.65%. "Sadness" is misclassified as "disgust" with 15.21% and "anger" is misclassified as "disgust" with 9.06%. This can be explained by the fact that these expressions have similar patterns, when the nose and mouth region are not taken into account. They include a strong lowering of eyebrows and a narrowing of the eyes. Furthermore, "anger" and "disgust" can both imply a contraction of the jaw muscles.
Both, "neutral" and "surprise", are misclassified with each other at about 10%. This may be due to the fact, that no electrodes of the subset are located close to the mouth and in this way small mouth openings during "surprise" may not be picked up.

		Prediction [%]						
		Anger	**Disgust**	**Fear**	**Happ.**	**Neutral**	**Sadn.**	**Surpr.**
Target	**Anger**	80.52	9.06	3.02	0.00	1.35	3.02	3.02
	Disgust	25.42	57.08	3.02	0.00	0.00	13.65	0.83
	Fear	2.29	2.29	81.25	0.00	1.67	5.42	7.08
	Happiness	0.00	0.00	0.10	98.54	0.00	0.42	0.94
	Neutral	0.00	0.00	0.31	0.00	89.27	0.00	10.42
	Sadness	2.19	15.21	3.96	0.10	0.63	77.71	0.21
	Surprise	0.31	0.00	5.83	0.83	9.58	0.00	83.44

Table 6.5: Average confusion matrix of channel subset [1 2 5 6].

The positions 1, 5 and 6 have in common that they all measure facial areas with a minimum of underlying muscles. Therefore, the crosstalk of other muscles at these positions is lower as compared to those closer to the facial center.

The positions with the highest impact on the recognition result are the m. zygomaticus major (channel 5) and m. frontalis (channel 1). The subset of this two positions still provides an ERR of 76.07% and a subject average VRR of 91.35% for "happiness" and 84.37% for "neutral", whereas the other expressions score between 65.31% to 47.81%.

6.1.8 Conclusions

The evaluation showed that our system is capable of recognizing the 7 basic facial expressions included in the data corpus with a high accuracy. The best feature set is TD-Slope-Waveform. It was determined that a signal length of 400 ms provides the best classification results.
Furthermore, the main confusion between expression exists between "disgust" and "sadness", whereas "happiness" and "neutral" achieved the best correct classifications through out all evaluation tasks.

We showed that for the same subject it is possible to train a classification system, which can be reused on data recorded in another recording session, but with a decrease in ERR.
The sequential backward channel reduction allowed us to minimize the electrode placement to 4 positions by a ERR decrease smaller than 5%. Thereby, mainly those positions remained, which are placed about regions with a minimum of underlying muscles. When reducing the placement to 2 positions, over m. frontalis and m. zygomaticus major, still a system can be trained, which distinguishes between positive ("happiness"), "neutral" and negative expressions ("sadness", "fear", "anger").

6.2 Live Expression Recognition Demonstration

Based on the previously described results, a demonstration system was developed, capable of classifying posed facial expressions and displaying them on an avatar. As an avatar system, we used the Head Zero+ of the Thinking Head (TH) project[Pro09]. This avatar represents a head of the performance artist Stelarc [Ste09]. We built two prototype systems. The first system simulates a live recognition on the training and test data used in the evaluation. The second system is a live demonstration system that presents the capability of also displaying different expressions than the previously used. The following sections will summarize the general structure of these two systems.

6.2.1 Thinking Head

The 3D avatar we used was developed by the Thinking Head (TH) project, which is located at and administered by the MARCS Auditory Laboratories at the University of Western Sydney. The goal of TH is to build a head embodying human attributes to improve Human-Computer-Interaction. Furthermore, it is a research

platform to test software in a realtime environment. The system TH Head Zero+ (THZ+) includes communication modes for natural speech and keyboard input. The system is able to form and answers to questions, which are integrated in an a priori specified knowledge base. The visual abilities are to transform facial parts around the mouth according to a speech output. Other facial parts can be transformed by linearly combining 16 different predefined shapes. Up to two combinations can be merged simultaneously. Hence, the degree of detailed facial expressions is limited. The length and intensity of a specific displayed expression has to be defined before execution. These conditions were considered in the design of the live system.

6.2.2 Simulation System

We created a system simulating a live recording capable to recognize and display facial mimics on a 3D avatar head. For this, several modules were developed. Two separate processes were implemented in MATLAB: the first streams segments of the EMG raw data and the second module processes and classifies these signal segments. The classification result then is transfered to a Thinking Head module to be displayed on the avatar.

In the first phase, a subset of the data is used in the training phase (figure 6.6). The training set is divided into a training and optimization part. The optimization is performed analog to the optimization task described in the previous evaluation sections.

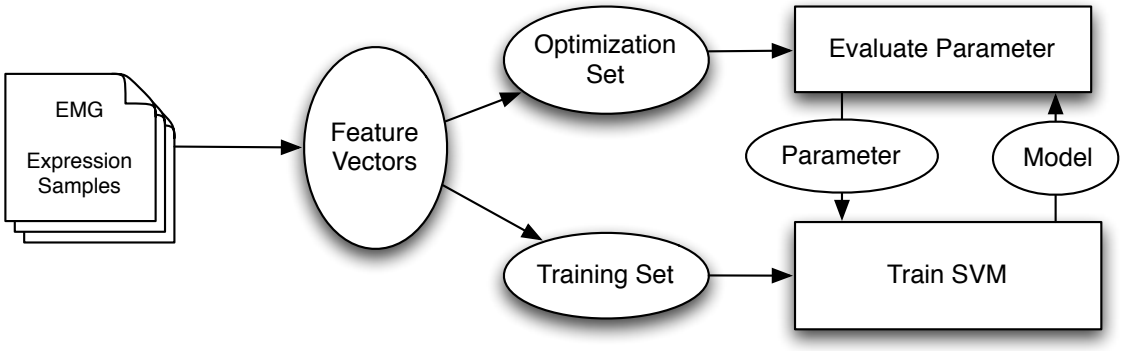

Figure 6.6: Training phase of the SVM model

After the training phase, the live recognition system simulation is started, where the remaining data is sequentially writing signal segments into an external file. This file is imported and the segments are then processed by the recognition module and classified with the beforehand trained SVM model.

The THZ+ offers constrained possibilities to emote facial expressions. Therefore, we limited the expression recognition by sending evocation commands at a maximum rate of 2 seconds only. This allowed to implement a further recognition result optimization by including a majority vote over sequential feature vector classification results, analog to the system used in the previous evaluation.

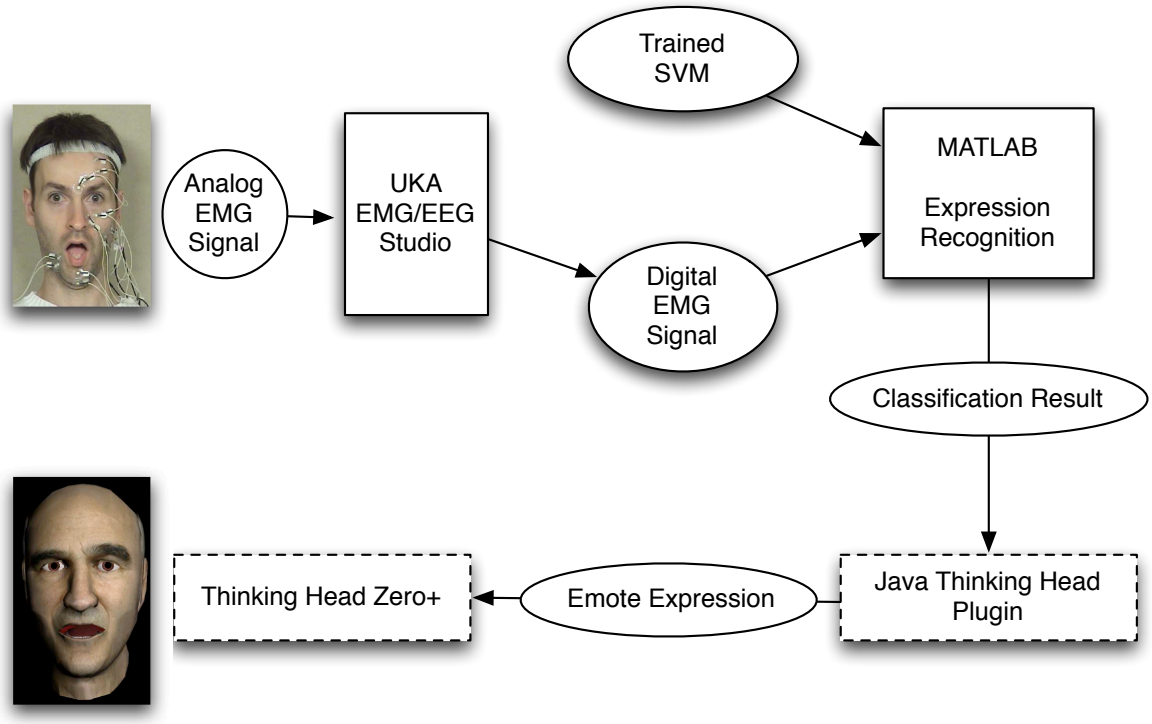

Figure 6.7: Demonstration phase: The analog EMG signal is analyzed for channel activation and the expression is recognized. The recognition result is sent to a TH plugin and the expression is emoted on the TH 3D head.

6.2.3 Live Demonstration System

The demonstration system classifies a facial expressions posed by the user and displays it on a 3D head. The type of facial expressions used in the demonstration is limited by the predefined expressions included in the THZ+ and classification possibilities through the presented method.

An example implementation included the expressions ”Smile”, ”Anger”, ”Wonder”, ”Wink” and ”Neutral” and the four electrode positions including m. frontalis, m. corrungator, m. zygomaticus major and m. labii superioris. To create a set of training data, we use a similar setup as employed for the data collection. After the electrodes are attached on the prepared skin, the system has to be trained with examples for each expression. For this, the user poses all expressions that are included in the demonstration expression set. In this system, the name of each expression displayed on the screen is used as the stimulus to pose each expression during the training, e.g., ”Smile”. The training phase uses a push-to-pose method. Here, the user is indicating the beginning and end of an expression by pressing a button. The expressions are displayed in a random order to rule out training effects. After a successful recording the expression samples, the training phase is started according to section 6.2.2 (Fig. 6.6). The SVM model training process is analog to the one used in the live simulation system.

Subsequently, the live demonstration phase is started (Figure 6.7). The user poses a random expression included in the training set and indicates the expression start and stop via the push-to-pose mode. The signal stream of the EMG equipment is recorded via the the UKA EMG/EEG Studio. The recognition system imports the digitized EMG signal, preprocesses this data and employs the beforehand trained SVM model. Then, the classified expression of the recorded sample is sent to the Thinking Head Module. Finally, the expression is displayed on the 3D head.

7. Conclusion

In this thesis, we developed a facial expression recognition system based on facial surface EMG. An appropriate electrode placement had to be determined and a data corpus was built. The corpus includes 7 posed facial expressions, including anger, disgust, fear, happiness, neutral, sadness and surprise. In each recording session 50 examples of each of these expression were recorded.

A system was developed, which extracts features of a digitized EMG signal and classifies facial expressions based on a trained SVM model. We compared different feature extraction methods in the time domain. The best results were achieved by a combination of features used in sEMG based prothesis control and silent speech recognition. The developed system reached a subject average expression recognition rate (ERR) of 92.02% in session and subject dependent classification tasks. Also, a session independent task could be performed with an average ERR of 86.79%.

Furthermore, we demonstrated how the number of electrodes can be reduced by 50% to 4 positions with an average recognition rate of 87.86% by an ERR decrease of 4.16%. This subset includes only positions on the forehead and outer cheek region. One advantage of this reduction is the possibility to use equipment with fewer channels. The main advantage is that these positions do not require the placement of electrodes close to eyes, mouth or nose and further no cables have to cross the face. In this way, a system can be created, which could be integrated into an every day device, such as an extension to a headset for mobile communication, which are already widely accepted and used.

We developed a live demonstration system to recognize facial expressions and displaying them on a 3D avatar. This can be implemented into communication processes using virtual environments. This system can be extended and tested to include more facial expressions. This can lead to an application in voluntarily device control through mimic muscles. For quadriplegic patients, this would provide an enormous advancement to their independence.

There are many research ideas for future work to enhance the system. Below, some research ideas are proposed.

Session Independence

As we have shown, there was a significant recognition rate of expressions posed by the same person over two sessions, including a complete removal and reattachment of the electrodes. To confirm these findings, several sessions over longer time periods and several different subjects should be analyzed. The research task is to find ways to create a reusable recognition system. To test this, more subjects and sessions should be recorded to gather more reliable experimental results.

Detailed Expression Recognition

A huge advancement of the system is the development of a recognition system capable to determine the intensity of a specific type of facial expressions. A major advantage of EMG could be used in the detection and processing of micro muscle activities that do not result in visible facial movements. Furthermore, more facial expressions can be analyzed to test the limits of the given system.

Electrode Placement

To minimize and optimize classification performance further, the electrode placement at different electrode positions could be analyzed and then be minimized in the described way. The application of electrode arrays including more than two electrodes per position could yield a more detailed detection of muscle activities and allow the detection and removal of crosstalk. This may be done, as crosstalk signals will not occur at all detection positions in the same strength and in this way can be distinguished from an activation of an underlying muscle.

A. Evaluation Result Tables

In this part of the appendix detailed experimental results are listed. Table A.1 shows the majority decision based expression sample recognition rates (ERR) and Table A.2 shows the feature vector based recognition rate (VRR) for TD0, MimikFeat1 and MimikFeat 2 for different window sizes. Table A.3 and A.4 present detailed average recognition results for the analyzed TD variants and different context lengths for different window sizes. The used system was trained on the feature vectors of 30 training samples for each expression and tested on the feature vectors of 20 different samples. The dimensionality of the feature space was reduced to $k - 1$ dimensions through a LDA with $k = 7$, the number of used expression classes (anger, disgust, fear, happiness, neutral, sadness and surprise). Maximum recognition results are marked bold.

	Expression Recognition Rate (ERR) [%]								
Win/Shift [ms]	**25/ 13**	**50/ 25**	**75/ 35**	**100/ 50**	**150/ 75**	**200/ 100**	**300/ 150**	**400/ 200**	**500/ 250**
	A								
TD0	91.43	92.86	93.57	94.29	94.29	93.57	93.57	95.00	95.71
MimikFeat1	91.43	91.43	92.14	92.14	90.71	91.43	90.71	91.43	90.00
MimikFeat2	92.86	94.29	94.29	95.71	93.57	92.86	93.57	90.71	91.43
	B1								
TD0	80.71	81.43	84.29	82.86	84.29	85.71	85.00	85.00	85.00
MimikFeat1	77.86	80.00	82.14	82.86	82.86	85.00	84.29	85.00	85.00
MimikFeat2	80.00	80.71	82.14	82.14	82.86	81.43	87.14	84.29	82.14
	B2								
TD0	84.29	85.00	85.00	85.00	85.00	85.71	87.86	87.14	87.14
MimikFeat1	83.57	83.57	83.57	85.00	84.29	83.57	82.14	81.43	83.57
MimikFeat2	85.71	84.29	84.29	86.43	85.71	86.43	87.86	86.43	86.43
	C								
TD0	86.43	87.86	88.57	89.29	89.29	92.14	89.29	90.00	90.00
MimikFeat1	87.86	90.71	90.00	90.00	90.71	90.00	90.71	88.57	85.71
MimikFeat2	87.86	90.71	90.71	91.43	92.14	90.00	91.43	86.43	87.86
	Subject Average ERR								
TD0	85.71	86.79	87.86	87.86	88.21	89.29	88.93	89.29	**89.46**
MimikFeat1	86.67	87.98	88.33	**88.69**	88.33	88.57	88.21	87.74	86.67
MimikFeat2	87.86	89.17	89.40	90.48	90.00	88.93	**90.83**	87.50	87.86

Table A.1: Expression Sample Recognition Rates (EER) for TD0, MimikFeat1, MimikFeat2

	Feature Vector Based Recognition Rate (VRR) Results [%]								
Win/Shift [ms]	**25/ 13**	**50/ 25**	**75/ 35**	**100/ 50**	**150/ 75**	**200/ 100**	**300/ 150**	**400/ 200**	**500/ 250**
	A								
TD 0	79.96	85.48	87.66	88.99	90.46	90.83	**91.82**	91.25	91.79
MimikFeat1	79.27	84.10	85.83	86.67	87.35	87.48	88.29	87.86	**88.57**
MimikFeat2	80.78	85.03	86.31	87.71	88.66	89.03	**89.21**	88.37	87.86
	B1								
TD 0	71.06	74.75	75.94	76.56	77.91	79.25	80.65	81.16	**81.19**
MimikFeat1	70.93	73.13	74.62	75.41	77.05	78.57	79.71	80.82	**81.00**
MimikFeat2	70.5	73.47	74.30	75.23	76.82	77.98	80.07	**81.12**	80.57
	B2								
TD 0	75.06	79.19	80.09	81.23	81.86	82.42	82.99	83.93	**84.05**
MimikFeat1	74.72	78.07	78.76	79.61	80.30	80.13	80.21	81.53	**81.57**
MimikFeat2	75.42	77.99	79.14	79.72	79.94	80.13	81.21	81.63	**82.00**
	C								
TD 0	74.72	79.41	81.36	82.60	83.80	**85.79**	84.29	84.73	83.93
MimikFeat1	75.19	80.10	81.18	81.41	81.64	82.23	**82.86**	82.24	82.57
MimikFeat2	76.08	80.17	81.24	81.58	82.68	83.07	83.00	83.37	**83.86**
	Subject Average VRR								
TD0	75.91	80.62	82.34	83.5	84.71	85.82	85.97	**86.18**	86.11
MimikFeat1	75.76	79.93	81.23	81.86	82.55	83.02	83.70	83.76	**84.14**
MimikFeat2	76.61	80.31	81.42	82.25	83.24	83.72	84.29	**84.37**	84.33

Table A.2: Feature Vector Based Recognition Rate (VRR) for TD0, MimikFeat1, MimikFeat2

	Average Majority Decision Expression Sample Recognition Rate (ERR) [%]								
Win/Shift [ms]	**25/ 13**	**50/ 25**	**75/ 35**	**100/ 50**	**150/ 75**	**200/ 100**	**300/ 150**	**400/ 200**	**500/ 250**
TD 0	85.71	86.79	87.86	87.86	88.21	89.29	88.93	89.29	**89.46**
TD1	86.79	87.5	88.39	88.04	89.11	**89.82**	88.21	88.93	88.75
TD2	87.5	87.86	88.39	87.86	**89.11**	88.75	85.18	83.93	-
TD3	88.04	88.04	87.68	88.04	**88.93**	88.04	78.93	-	-
TD4	88.04	**88.75**	88.21	88.57	88.57	86.61	-	-	-
TD5	88.21	88.39	88.39	**89.46**	86.79	83.04	-	-	-
TD0_slope	87.62	88.69	89.40	89.29	89.52	**91.55**	89.88	91.07	91.55
TD1_slope	88.45	89.88	89.88	90.36	91.07	**91.31**	89.17	90.12	89.05
TD2_slope	88.81	90.12	**90.83**	89.88	90.00	89.76	86.19	86.55	-
TD3_slope	90.12	**90.83**	90.48	90.36	89.40	87.98	79.76	-	-
TD4_slope	**90.12**	89.76	89.64	89.76	88.93	86.90	-	-	-
TD5_slope	**89.76**	89.29	89.64	89.29	88.21	83.69	-	-	-
TD0_waveform	88.21	88.81	89.64	89.52	90.00	**91.90**	90.95	91.19	91.07
TD1_waveform	88.33	89.52	90.48	90.60	90.83	**92.02**	90.00	90.36	90.00
TD2_waveform	88.93	89.29	90.83	90.48	**91.07**	90.71	87.62	84.17	-
TD3_waveform	89.29	**90.48**	90.48	90.24	89.88	90.48	76.31	-	-
TD4_waveform	89.17	**90.71**	90.36	90.71	89.29	88.21	-	-	-
TD5_waveform	89.76	**91.07**	90.71	89.88	88.69	81.19	-	-	-
TD0_slope_waveform	88.10	88.93	90.12	89.76	90.71	91.90	91.43	**92.02**	92.02
TD1_slope_waveform	88.45	89.52	90.60	90.71	91.55	**91.19**	90.36	90.95	88.81
TD2_slope_waveform	89.17	90.00	90.71	**90.95**	90.95	90.36	87.02	82.98	-
TD3_slope_waveform	89.40	90.48	**91.31**	90.95	89.52	89.88	75.36	-	-
TD4_slope_waveform	90.00	**91.07**	90.95	90.60	88.93	86.90	-	-	-
TD5_slope_waveform	90.00	**91.19**	91.07	90.24	87.98	79.76	-	-	-

Table A.3: Average Majority Decision Expression Sample Recognition Rate (ERR) of TD Variants

	Average Feature Vector Based Recognition Rate (VRR) for TD Variants [%]								
Win/Shift [ms]	**25/ 13**	**50/ 25**	**75/ 35**	**100/ 50**	**150/ 75**	**200/ 100**	**300/ 150**	**400/ 200**	**500/ 250**
TD 0	75.20	79.71	81.26	82.34	83.51	84.57	84.94	**85.27**	85.24
TD1	78.73	82.32	83.19	83.94	**85.36**	85.22	85.20	84.38	83.04
TD2	80.56	83.17	83.96	84.4	**84.99**	84.85	82.81	78.53	-
TD3	81.72	83.46	84.24	84.49	**84.67**	83.72	75.29	-	-
TD4	82.32	84.30	84.27	84.67	**84.79**	81.00	-	-	-
TD5	82.69	84.51	84.47	**85.04**	82.86	77.41	-	-	-
TD0_slope	75.4	80.02	81.84	82.52	83.86	84.86	85.57	**85.63**	85.57
TD1_slope	78.79	82.88	83.5	84.23	**85.37**	84.96	85.04	84.23	82.37
TD2_slope	80.7	83.75	84.84	84.59	**84.86**	84.45	82.53	77.59	-
TD3_slope	81.88	84.43	84.20	**84.82**	84.59	83.36	74.57	-	-
TD4_slope	82.44	84.04	84.33	**85**	84.7	80.41	-	-	-
TD5_slope	82.84	84.02	**84.73**	84.64	82.62	75.38	-	-	-
TD0_waveform	76.01	80.56	82.25	83.26	84.41	85.51	86.20	**86.65**	86.37
TD1_waveform	79.55	83.13	84.13	84.7	85.67	**86.1**	85.30	85.09	82.95
TD2_waveform	81.20	83.93	84.89	85.36	**86.08**	85.61	82.55	75.22	-
TD3_waveform	81.99	84.44	85.28	85.4	**85.46**	84.58	71.14	-	-
TD4_waveform	82.55	84.82	85.4	**85.48**	84.22	81.14	-	-	-
TD5_waveform	82.9	84.99	**85.37**	84.96	83.05	72.92	-	-	-
TD0_slope_waveform	76.15	80.71	82.62	83.51	84.76	85.86	86.12	86.58	**86.61**
TD1_slope_waveform	79.62	83.08	84.34	84.86	85.43	**85.78**	85.08	84.94	81.74
TD2_slope_waveform	81.30	84.10	85.08	85.39	**85.74**	84.97	81.73	74.02	-
TD3_slope_waveform	82.27	84.64	85.12	**85.58**	85.01	84.23	69.07	-	-
TD4_slope_waveform	82.7	84.9	85.31	**85.48**	84.08	80.00	-	-	-
TD5_slope_waveform	83.19	84.92	**85.66**	85.18	82.63	70.87	-	-	-

Table A.4: Average Feature Vector Based Recognition Rate (VRR) of TD Variants

Window size [ms]	**25**	**50**	**75**	**100**	**150**	**200**	**300**	**400**	**500**
Shift size [ms]	**13**	**25**	**35**	**50**	**75**	**100**	**150**	**200**	**250**
TD 0	25	50	75	100	150	200	300	400	500
TD 1	51	100	155	200	300	400	600	800	1000
TD 2	77	150	215	300	450	600	900	1200	1500
TD 3	103	200	285	400	600	800	1200	1600	2000
TD 4	129	250	355	500	750	1000	1500	2000	2500
TD 5	155	300	425	600	900	1200	1800	2400	3000

Table A.5: Time in ms to produce one TD Feature Vector in relation of different Context Length and Window Sizes

Bibliography

[ABB+04] L. B. P. Ang, E. F. Belen, Jr. Bernardo, R. A., E. R. Boongaling, G. H. Briones, and J. B. Coronel. Facial expression recognition through pattern analysis of facial muscle movements utilizing electromyogram sensors. In *TENCON IEEE Region 10 Conference*, volume 3, pages 600–603, Nov. 2004.

[And07] H. L. Andreassi. *Psychophysiology - Human Behavior and Physiological Response*. Lawrence Erlbaum Associates, Inc., fifth edition, 2007.

[BGV92] B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pages 144–152. ACM Press, 1992.

[BLP01] P. D. Boll, A. Lang, and R. F. Potter. The effects of message valence and listener arousal on attention, memory and facial muscular responses to radio advertisements. *Communication Research*, 28, 2001.

[Bur98] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2:121–167, 1998.

[CAG08] Z. Y. Chin, K. K. Ang, and C. Guan. Multiclass voluntary facial expression classification based on filter bank common spatial pattern. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 1005–1008, Aug. 2008.

[CE05] Jeffrey Cohn and P. Ekman. Measuring facial action by manual coding, facial emg, and automatic facial image analysis. In J. A. Harrigan, R. Rosenthal, and K. Scherer, editors, *Handbook of nonverbal behavior research methods in the affective sciences*, pages 9–64. 2005.

[CEHL02] A. D. C. Chan, K. Englehart, B. Hudgins, and D. F. Lovely. Hidden markov model classification of myoelectric signals in speech. *IEEE Engineering in Medicine and Biology Magazine*, 21(5):143–146, 2002.

[CHCL03] Department Of Computer, Chih-Wei Hsu, C.-C. Chang, and C.-J. Lin. A practical guide to support vector classification, technical report, 2003.

[CHOA81] R. M. Carney, B. A. Hong, M. F. O'Connell, and H. Amado. Facial electromyography as a predictor of treatment outcome in depression. *The British Journal of Psychiatry*, 138:485–489, 1981.

[CL09] C. C. Chang and C. J. Lin. LIBSVM: a library for support vector machines. http://www.csie.ntu.edu.tw/~cjlin/libsvm, 2009.

[Dar72] C. Darwin. *The Expression of the Emotions in Man and Animals*. J. Murray, London, 1872.

[Dar98] C. Darwin. *The Expression of the Emotions in Man and Animals, With an Introduction, Afterword and Commentaries by Paul Ekman*. Oxford University Press, 3 edition, 1998.

[DHS01] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification*. Wiley, New York, 2. edition, 2001.

[Dim90] U. Dimberg. Facial electromyography and emotional reactions. *Psychophysiology*, 19:643–647, 1990.

[DL02] C. J. De Luca. Surface electromyography: Detection and recording. http://www.delsys.com/Attachments_pdf/WP_SEMGintro.pdf, 2002.

[DSH05] P. Deetjen, E.-J. Speckmann, and J. Hescheler. *Physiologie*. Urban & Fischer, München, 4. edition, 2005.

[DTE00] U. Dimberg, M. Thunberg, and K. Elmehed. Unconscious facial reactions to emotional facial expressions. *Psychological Science*, 11(1):86–89, 2000.

[EF71] P. Ekman and W. V. Friesen. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2):124 – 129, 1971.

[EF78] P. Ekman and W. V. Friesen. *The facial action coding system (FACS): A technique for the measurement of facial action*. Palo Alto, CA: Consulting Psychologists Press, 1978.

[EH03] K. Englehart and B. Hudgins. A robust, real-time control scheme for multifunction myoelectric control. *IEEE Transactions on Biomedical Engineering*, 50(7):848–854, 2003.

[Ekm78] P. Ekman. Facial signs: Facts, fantasies, and possibilities. *Sight, Sound and Sense*, pages 124–156, 1978.

[ER05] P. Ekman and E. L. Rosenberg. *What the Face Reveals*. Oxford University Press, 2 edition, 2005.

[FC86] A. J. Fridlund and J. T. Cacioppo. Guidelines for human electromyographic research. *Psychophysiology*, 23(5):567–89, 1986.

[FL03] B. Fasel and J. Luettin. Automatic facial expression analysis: A survey. *Pattern Recognition*, 36, 2003.

[FSF84] A. J. Fridlund, G. E. Schwartz, and S. C. Fowler. Pattern recognition of self-reported emotional state from multiple-site facial emg activity during affective imagery. *Psychophysiology*, 21(6):622–37, 1984.

[GKF95] P. Gosselin, G. Kirouac, and Y. François. Components and recognition of facial expression in the communication of emotion by actors. *Journal of Personality and Social Psychology*, 68(1):83 – 96, 1995.

[HE95] J. Hager and P. Ekman. The essential behavioral science of the face and gesture that computer scientists need to know. In *Proc. Int'l Workshop Automatic Face- and Gesture-Recognition*, pages 7–11, 1995.

[JKS05] H. Jeong, J.-S. Kim, and W.-H. Son. An emg-based mouse controller for a tetraplegic. In *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, volume 2, pages 1229–1234, Oct. 2005.

[JSW+06] S.-C. Jou, T. Schultz, M. Walliczek, F. Kraft, and A. Waibel. Towards continuous speech recognition using surface electromyography. In *Proc. Interspeech*, Pittsburgh, PA, Sep 2006.

[LH92] R. Lamb and D. Hobart. Anatomic and physiologic basis for surface electromyography. in selected topics in surface electromyography for use in the occupational setting: Expert perspective. u.s. department of health and human services. no 91-100. 1992.

[LL03] H.-T. Lin and C.-J. Lin. A study on sigmoid kernels for svm and the training of non-psd kernels by smo-type methods. Technical report, Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan, 2003. Available at http://www.csie.ntu.edu.tw/˜cjlin/papers/tanh.pdf.

[LNC03] J. T. Larsen, C. J. Norris, and J. T. Cacioppo. Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology*, 40:776–785, 2003.

[May05] C. Mayer. *UKA EMG/EEG Studio v2.0.*, 2005.

[Med03] Becker Meditec. http://www.becker-meditec.de, 2003.

[Meh68] A. Mehrabian. Communication without words. *Psychology Today*, 4:53–56, 1968.

[MHMSW05] L. Maier-Hein, F. Metze, T. Schultz, and A. Waibel. Session independent non-audible speech recognition using surface electromyography. 2005.

[Mik08] R. Mikut. *Data Mining in der Medizin und Medizintechnik*. Universitätsverlag, Karlsruhe, 2008.

[MMMW07] E. J. Moody, D. N. McIntosh, L.J. Mann, and K.R. Weisser. More than mere mimicry? the influence of emotion on rapid facial reactions to faces. *Emotion*, 7, 2007.

[NT94] K. Nagao and A. Takeuchi. Speech dialogue with facial displays: Multimodal human-computer conversation. In *ACL*, pages 102–109, 1994.

[Pro09] Thinking Head Project. http://thinkinghead.edu.au, October 2009.

[PZA+98] J. E. Pessa, V. P. Zadoo, E. K. Jr. Adrian, C. H. Yuan, J. Aydelotte, and J. R. Garza. Variability of the midfacial muscles: analysis of 50 hemifacial cadaver dissections. *Plastic and Reconstructive Surgery*, 102(6):1888–93, 1998.

[RHY06] M. B. Raez, M. S. Hussain, and Mohd F. Yasin. Techniques of emg signal analysis: detection, processing, classification and applications. *Biological procedures online*, 8:11–35, 2006.

[Rin84] W. E. Rinn. The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions. *Psychological Bulletin*, pages 52–77, 1984.

[SC01] K. L. Schmidt and J. F. Cohn. Human facial expressions as adaptations: Evolutionary questions in facial expression research. *Yearbook of Physical Anthropology*, 44:3–24, 2001.

[Sch06] R. Schandry. *Biologische Psychologie : ein Lehrbuch*. Beltz PVU. Beltz PVU, Weinheim, 2. edition, 2006.

[Sch08] K. Schaaf. Eeg based emotion recognition, September 2008.

[Sco03] D. Scott. Important factors in surface emg measurement. http://www.bortec.ca/Images/pdf/EMGmeasurementandrecording.pdf, 2003.

[SF09] O. Spindler and T. Fadrus. Grimace project. www.grimace-project.net, October 2009.

[SK05] K. Sagawa and O. Kimura. Control of robot manipulator using emg generated from face, 2005.

[Sob00] J. Sobotta. *Atlas der Anatomie des Menschen*, volume 1: Kopf, Hals, obere Extremität : 76 Tabellen. Urban und Schwarzenberg, München, 21. edition, 2000.

[SSSbKW06] M. Schünke, E. Schulte, U. Schumacher, and Design by K. Wesker. *Prometheus - Lernatlas der Anatomie*, volume [3]: Kopf und Neuroanatomie. Thieme Verlag, Stuttgart, New York, 2006.

[Ste09] Stelarc. http://www.stelarc.va.com.au, October 2009.

[vdMH08] L. J. P. van der Maaten and G. E. Hinton. Visualizing high-dimensional data using t-sne. *Journal of Machine Learning Research*, 9:2579–2605, November 2008.

[WS09] M. Wand and T. Schultz. Towards speaker-adaptive speech recognition based on surface electromyography. In *International Conference on Bio-inspired Systems and Signal Processing*, 2009.