Graz University of Technology

Diploma Thesis

Electrolarynx Control using Electromyographic Signals

conducted at the
Signal Processing and Speech Communications Laboratory
Graz University of Technology, Austria

by
Clemens Amon, 0431129

Supervisors:
DI Dr. techn. Martin Hagmüller
DI Anna Katharina Fuchs

Assessor:
DI Dr. techn. Martin Hagmüller

Graz, January 8, 2014

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

date (signature)

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am (Unterschrift)

Abstract

The larynx, which contains the vocal folds, is essential for the production of speech. If the larynx is surgically removed, typically due to cancer, the patient loses the ability to speak. One possibility to be able to communicate again is to use a hand-held, battery-driven device - the electrolarynx. The device is held against the neck and its vibrations are transmitted through the neck into the vocal tract. The patient can produce speech by changing the shape of the vocal tract in the oral cavity such as the tongue, the velum and the lips. A major drawback of the device is that it occupies the use of one hand, as the excitation signals has to be turned on and off for every spoken word or phrase. To overcome this problem, this work presents an electromyography-based approach to detect voiced speech.

An electromyography acquisition shield for a micro-controller board is introduced and different processing and classification methods are implemented and evaluated. Signal envelopes are calculated using root mean square, Hilbert transform and Teager energy operator algorithms. The on/off classification is implemented using single, double and adaptive threshold detection algorithms. Detection results are compared to the corresponding speech ground truth and time-dependent error rates such as front end error, back end error, middle speech error and noise detected as speech rates are calculated. The time-independent block detection ratio, an indicator of the likelihood of interruptions during activity and the time constant detection smoothing, a detection smoothing algorithm to avoid short interruptions in real-time electromyography detection are introduced. Configurations which offer best detection results are proposed and optimal user dependent operational points are determined using receiver operating characteristics analysis. It is shown that a combination of the Hilbert transform envelope and the time constant detection smoothed activity detection of the double threshold detection algorithm yields a minimum mean total error of 6,7% for real-time use.

Kurzfassung

Der Kehlkopf (lat. larynx) enthält die Stimmlippen und ist für die Sprachproduktion essentiell. Muss dieser als Konsequenz von Kehlkopfkrebs operativ entfernt werden, verliert der Patient die Fähigkeit zu sprechen. Eine Möglichkeit der Patientin/dem Patienten verbale Kommunikation zu ermöglichen, bietet der tragbare, batteriebetriebene Elektrolarynx. Dieses Gerät wird gegen den Hals gehalten und leitet Vibrationen in die Rachenhöhle. Die Patientin/der Patient kann durch die Veränderung der Form der Stimmorgane im Vokaltrakt wie Zunge, Gaumensegel und Lippen Sprache erzeugen. Ein großer Nachteil dieses Gerätes ist, dass beim Benutzen immer eine Hand benötigt wird, die das Anregungssignal für jedes Wort oder jede Phrase ein- und ausschaltet. Um diesem Problem entgegenzuwirken, wird in dieser Arbeit ein Ansatz präsentiert, in dem Muskelsignale, die bei stimmhafter Sprache auftreten, mittels Elektromyographie erkannt und zur Steuerung des Elektrolarynx eingesetzt werden.
Eine elektronische Schaltung zur Signalerfassung, basierend auf einem Mikro-Controller, wird vorgestellt und verschiedene Methoden zur Signalverarbeitung und Klassifizierung werden implementiert und evaluiert. Zur Berechnung der Hüllkurve des Signals werden folgende Methoden angewandt: Root Mean Square, Hilbert-Transformation und Teager Energy-Operator. Die Klassifizierung der Sprachaktivität wird mittels Single Threshold, Double Threshold und Adaptive Threshold Detection durchgeführt. Die Resultate der Aktivitätserkennung werden mit Referenzdaten der Sprache verglichen und es werden zeitabhängige Fehlerraten wie Front End Error, Back End Error, Middle Speech Error und Noise Detected as Speech berechnet. Die zeitunabhängige Block Detection Ratio, ein Maß für die Störanfälligkeit der Aktivitätserkennung, und das Time Constant Detection Smoothing, ein Algorithmus zur Glättung von kurzen Unterbrechungen in der Aktivitätserkennung für Echtzeit-Systeme, werden vorgestellt. Kombinationen, die zu bestmöglichen Resultaten führen, werden präsentiert. Mittels Receiver Operating Characteristics-Analyse werden Arbeitspunkte ermittelt, die optimale Ergebnisse für den jeweiligen Benutzer liefern. Eine Kombination aus Hilbert Transformation, Double Threshold Detection und Time Constant Detection Smoothing stellte sich als beste Variante heraus. Mit dieser Methode wird ein minimaler mittleren Fehler von 6,7% erzielt.

Acknowledgments

I would like to thank my supervisors DI Dr. techn. Martin Hagmüller and DI Anna Katharina Fuchs for their constant guidance and support during the research and writing of this thesis. Both of them were a valuable help. Cheers to my friends and fellow students for a great time in Graz, for their help and for so many discussions about my ideas, tasks and excitements. I owe my deepest gratitude to my parents for their unconditional support in my decisions, their patience and understanding throughout all my studies.

Contents

1 Introduction

1.1 Motivation

The electrolarynx (EL) offers one possibility for speech production for people, who have lost the ability to speak. It is easy to use and does not demand much training. The need of triggering the device by hand leads to limitations in the user's daily life, as one hand is permanently occupied while talking. Finding methods measuring muscle activity at the neck and process them to improve this device in terms of their usability is the major aim of this work.

1.2 Problem Definition and Limitations

Conventional EL systems have to be turned on and off using a button placed at the surface of the hand-held device. The purpose of this thesis is to design an easy-to-use electromyography (EMG) signal acquisition hardware and provide methods for a system in which neck muscle EMG activity is detected and used as input for the detection of word and sentence boundaries of EL speech. Using these methods will increase the ease of use in a way that the user doesn't have to use his hand to push the on/off button. A higher degree of freedom is the result of this free-hands EL system and will enable the user to interact with his environment in a more natural way.

A database of sentences, spoken by two healthy subjects, was recorded to develop and evaluate the proposed signal processing methods. It is assumed that muscular signals corresponding to disordered speech has the same structure as muscle signals from healthy speakers and are achieving similar results in the detection of on and off messages.

1.3 Organization of the Thesis

This document is divided into the following chapters:

Chapter 2 describes the background of speech production and the muscular system that are involved in this process. The EL system is introduced, as well as the principal operation mode of surface electromyography (sEMG) is explained. Related literature dealing with the idea of using sEMG for the enhancement of disordered speech is presented.

Chapter 3 explains the hardware development process and the hardware system design. The single parts of the electrical circuit and the simulation of it are described. Moreover the recorded speech database is presented.

Chapter 4 presents the proposed signal processing methods for detection. Different features extracted from the recorded and preprocessed muscle signals are classified by different detection algorithms. A system to calculate time-dependent and time-independent error rates is presented.

Chapter 5 provides the evaluation of proposed combinations of methods. Error rate distributions are presented and analyzed. Optimal working points are determined using receiver operating characteristics (ROC) analysis. Algorithms that offer detection at the lowest possible error rate and their characteristics as well as their parameters are discussed in this chapter.

In **Chapter 6** the conclusions of the work are presented based on the evaluation in chapter 5. Future work and an outlook to the implementation of the proposed system is presented.

In the **Appendix** the hardware system circuit and the design of the printed circuit board (PCB) as well as the changelog are provided. The whole database, a set of 114 sentences that were recorded by two speakers, is listed.

2 Background

2.1 Anatomy and Physiology

In this section the anatomic and physiologic basics of the humans vocal tract are described. This is necessary for understanding the concepts of the presented work.
The main part in human speech production is the vocal tract which consists of the vocal organs presented in figure 2.1. An air stream is produced when exhale and forced through the glottis surrounded by the vocal chords. Position and tension of the vocal chords are varying the air stream. With this air stream voiced (periodic manipulation) and unvoiced (no or slow manipulation of the air stream) sounds can be produced. The larynx, also called voice box, houses the vocal chords and is responsible to avoid aspiration of food into the trachea [21]. Due to cancer it might become necessary to remove the larynx surgically (total laryngectomy). This results in the inability to speak in a normal way and the patient has to breath through the stoma placed in the neck. The stoma is a hole which serves as an airway for total laryngectomees. The connection between the trachea and the pharynx (throat) is disrupted. Thus, the production of voiced speech is not possible. Most unvoiced consonants can still be produced in the oral cavity.

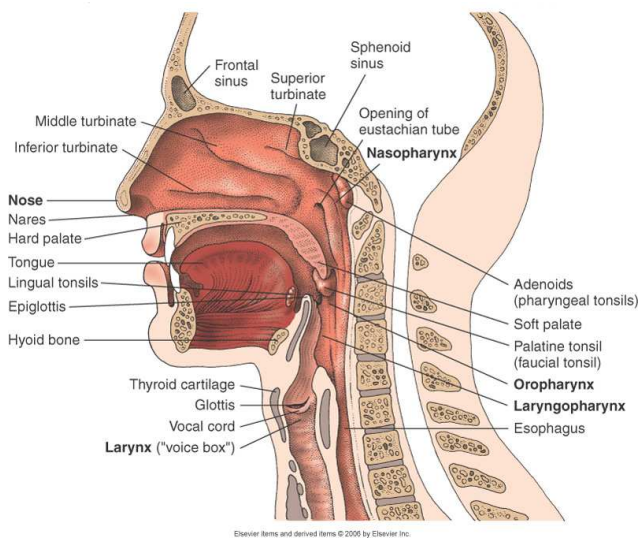


Figure 2.1: Anatomy of the human vocal tract.

2.2 Electrolarynx

A common way to reobtain speech is to use a handheld, battery-driven device, the electrolarynx (EL). Held against the neck, an excitation signal is transmitted through the neck tissue which results in a sound wave inside the oral cavity. Changing the volume and the form of the oral cavity causes a variation in the formant structure of the excitation signal. Combining these manipulations of the excitation signal with unvoiced sounds that can be produced without air stream an intelligible EL voice can be produced. Different types of conventional EL are presented in figure 2.2.
The two mayor drawbacks of EL systems are the monotonic and indistinct voice. This is due to the constant fundamental frequency of the excitation signal. An additional drawback is that the user needs a hand to hold the device against the neck and turn the excitation on and off.

Figure 2.2: Different types of EL.

2.3 sEMG Measurement at the Neck

As mentioned in section 2.2, a drawback of conventional EL systems is the occupation of one hand of the user. To overcome this drawback, the proposal of this work is to use sEMG to improve the usability of the device by detecting speech activity which heads to hands-free interaction.
sEMG is the most common non-invasive method used in medical diagnostics to measure motor unit action potentials (MUAPs) and accordingly the superposition of them. Discrete motor units and their corresponding muscle fibers are activated by the nervous system [4]. A different method, but an invasive one and therefor not considered in this work, is the use of needle electrodes. The measured superposed MUAPs are the combination of the muscle fiber action potentials from all the muscle fibers of a single motor unit. In general EMG serves as a good indicator for the force produced by skeletal muscles. The frequency range of the signal extends between 1 Hz and 1000 Hz, while most of the energy is distributed from 20 Hz to 200 Hz [2].

A simple mathematical model of the raw EMG signal is presented in equation 2.1:

$$x(k) = \sum_{r=0}^{N-1} h(r)e(k-r) + w(k) \tag{2.1}$$

where $x(k)$ is the EMG signal at time k, $e(k)$ are the impulses, Dirac delta functions, of the individual motor units, r represents the time delay of each impulse, $h(r)$ is the impulse response of the MUAP, $w(k)$ is the zero mean additive white Gaussian noise, and N is the number of motor unit firings [28]. This model shows the response of the system, here the MUAP, to a series of delayed impulses and does not describe the filtering characteristics of the tissue. Other

mathematical representations of the EMG signal are presented and compared in [28]. In Figure 2.3, a recorded EMG signal and the corresponding EL speech signal are shown. The EMG signal is amplified and recorded using a conventional audio interface at a sampling frequency of 16 kHz. The EL speech signal is recorded simultaneously using a microphone headset. It can be seen that during speech activity, the EMG activity is increasing. Furthermore, a rise in the muscle activity can be seen before speech is recorded. This is caused by the pre-activation of the muscles as a preparation to start speaking. This was first shown in the studies of J. Atkinson in 1978 [3]. In this recording a short burst of muscle activity with small amplitude can be seen at $t = 1s$. As no associated EL speech activity (green) can be seen, this may arise from swallowing or due to a short movement to prepare speaking.

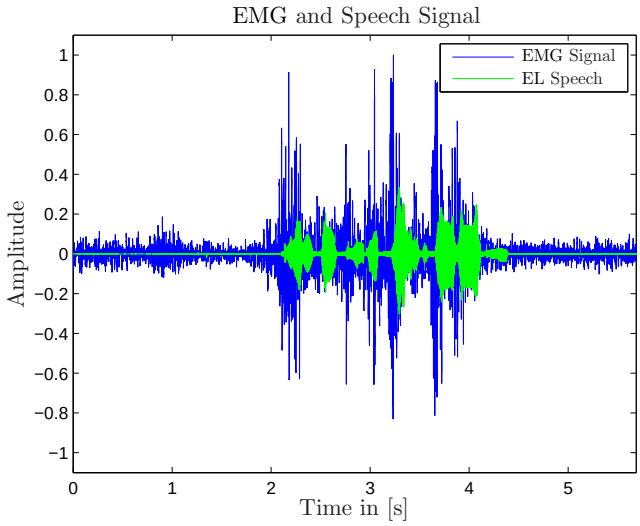


Figure 2.3: Recorded healthy male speaker saying the German sentence "Opa fährt ein blaues Fahrrad". sEMG signal with electrodes placed at the neck compared to the recorded EL speech signal.

2.4 Related Work

In 2004, Goldstein et al. [11] used surface electrodes to detect electric signals from neck muscles. Using single threshold detection (STD), described in section 4.5.1, on/off messages were fed to a hands-free EL held against the neck by a brace. Low-pitch vowels were found to generate maximum EMG activity from neck strap muscles.

This approach was improved by Kubert et al. [19, 29] in 2009. EMG envelopes produced by an analog circuit (low-pass filtered with cut off frequencies of 1 and 5 Hz) were used to generate on/off messages and a pitch contour fed to the EL. A "slow" EMG amplitude (smoothed by a 1 Hz low-pass filter) controls the fundamental frequency of the excitation signal of the EL and a "fast" 5 Hz low-pass filter amplitude is used to control activation and termination of the excitation. For the activity detection double threshold detection (DTD), described in section 4.5.2, was used. There the offset threshold is an adjustable ratio of the onset threshold. Users are able to adjust this ratio manually to tune the detector according to different optimal criteria, like inter-session (e.g. sensor position, skin preparation, etc.) and inter-subject differences (e.g. subject dependent signal-to-noise ratio (SNR), pre-activation time, skin conductivity, etc.).

Another approach of exciting the tissue to transmit sound energy is presented in [26]. An artificial larynx consisting of a piezo-sounder connected to a flexible plastic tube is controlled by

neck EMG. The plastic tube is inserted into the stoma and is transmitting the excitation signal into the oral cavity. On/off information, classified using STD, as well as pitch information is extracted from the EMG signal.

A wireless version of the sEMG-controlled EL is introduced in [13]. A small device housing the sensors, the amplification and digital conversion sections, as well as a radio frequency (RF) transmitter is placed at the neck and sends on/off and pitch messages to the handheld EL. The user can operate the device like a typical EL with the on/off button or supplement or gate the sEMG-based control. It is also possible to deactivate the button and use the device automatically controlled via sEMG activity as a hands-free system.

In 2011, Nakamura et al. [17,23] introduced an approach where a support vector machine (SVM) was used to recognize individual frames in the recorded EMG data as unvoiced and voiced. Inside the voiced frames the fundamental frequency was estimated using a Gaussian mixture Model (GMM)-based voice conversion. The aim of this work is EMG-recognition and in the next step EMG-to-speech conversion. The EMG signal was a analog high-pass filtered five channel EMG sampled at 600 Hz. Best results were achieved by delaying the controlling signal for activation and pitch contour by 50 ms. This time period was found to be the mean latency of speech compared to the recorded EMG signal. This work is based on a fundamental work on EMG-based recognition of silent speech by Janke in 2010 [16]. Presented methods and the voiced/unvoiced information in terms of EL can be used to control the on/off button of the EL. The estimated pitch contour was used to regulate the fundamental frequency to overcome the problem of the monotony of conventional EL hand-helds.

3 Signal Acquisition

3.1 EMG Hardware System

The developed sEMG hardware system consists of three parts. The sensor strap which is presented in section 3.1.1, the ARDUINO©DUE compatible bio-signal shield and an ARDUINO©DUE micro-controller board which serves as a host for the shield and the connected strap. The aim of the design is the development of a small, battery-operated, real-time bio-signal acquisition system. A schematic overview of the hardware system is shown in the block diagram in figure 3.1.

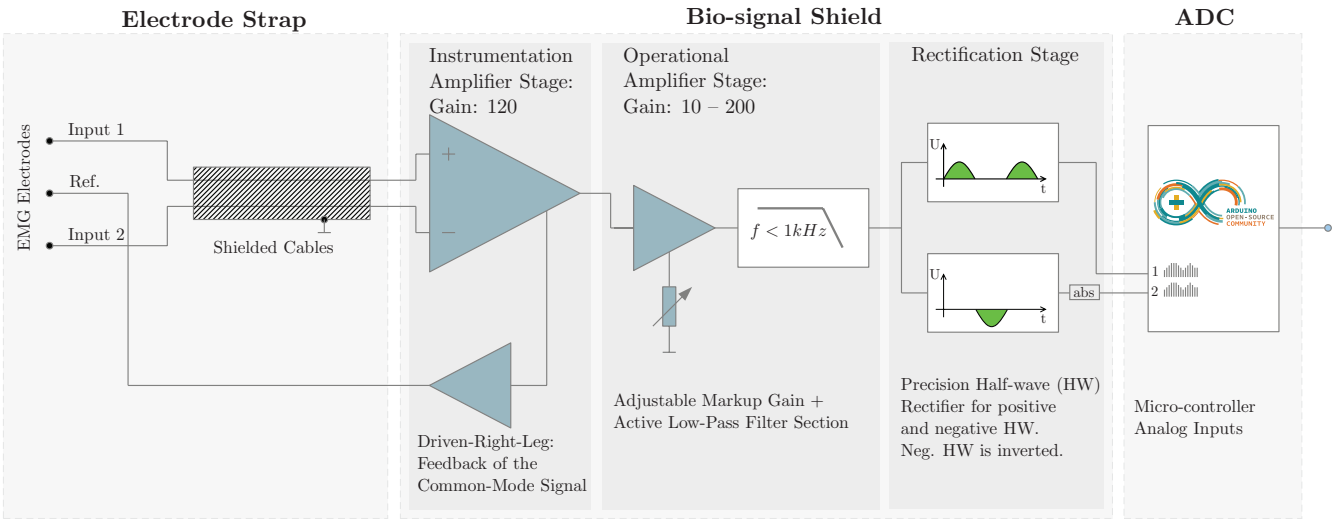


Figure 3.1: Schematic overview: Block diagram of the developed hardware system consisting of the electrode strap, the ARDUINO©DUE compatible bio-signal shield and the micro-controller which serves as a host.

The strap holds three electrodes to detect sEMG signals from the neck and is connected to the input of an instrumentation amplifier located on the bio-signal shield. A driven-right-leg (DRL) circuit feeds the signal back to the body to improve the common-mode rejection ratio (CMRR). The CMRR is an indicator of how well two similar signals can be suppressed by

a differential amplifier. The output of the instrumentation amplifier is connected to the input of an operational amplifier. The gain can be manually modified to adapt the amplification to the current conditions. After a low-pass filter where higher frequency noise is filtered out, the positive and negative half-wave are split and fed (the positive half-wave as it is, the negative half-wave inverted to become positive) to two discrete analog inputs of the micro-controller. Using this method, a higher resolution (bit rate) of the digitized signal amplitude can be achieved. After the analog to digital conversion, the values of the negative signal are subtracted from the values of the positive half-wave and the digital signal is available in the resolution of two times the resolution of one input channel. The ARDUINO©DUE has a maximum bit rate of 12 bit per input channel. As a result, the digitized sEMG signal has a theoretical amplitude resolution of 13 bit. The design and function of the bio-signal shield as well as the role of the micro-controller are presented in detail in section 3.1.2.

3.1.1 EMG Electrodes Strap

The sensor strap contains the three sEMG electrodes necessary to detect the EMG signal. The strap is designed to be worn around the neck. This ensures correct electrode positions. The electrodes are "Bluetrodes" Ag/AgCl (silver/silver-chloride) reusable electrodes by BETTERWITS®. Compared to self-adhesive electrodes, this reusable electrodes are less noisy and therefor provide a higher SNR. Other sensor systems with integrated electronics like the parallel-bar EMG sensors by DelSys®are expected to work very reliable as well and provide very good, low-noise signals. Due to their relatively high costs (150 €/piece) compared to self-adhesive ($\sim$ 0.50 €/piece) and reusable Ag/AgCl electrodes ($\sim$ 3 €/piece), this types of electrodes are not used or evaluated in this work. For the acquisition, two sEMG signals are detected by two electrodes placed 10 mm lateral (farther from the middle) to the ventral (pertaining to the front) midline of the neck at the location below the chin surface [24]. This signals are fed to the bio-signal shield and amplified and processed there. The sensor location was selected considering the study of Stepp et al. [29] and depending on the characteristics of the used strap. A third electrode serves as a reference electrode. Using this connection to the body, the amplifier is feeding back an inverted error signal to reduce the unwanted common mode. The cables connecting the strap to the bio-signal shield have to be shielded to reduce influences by artifacts caused by cable movement and contact to cables. A useful and simple way to fit the requirements to the cable is the use of a universal serial bus (USB) cable. This cable includes up to four shielded wires and therefore is perfect in this context.
To use this strap in an optimal manner, the skin has to be prepared prior to electrode placement using either a specified skin preparation gel or by cleaning the skin with alcohol or similar fluids.

3.1.2 Bio-Signal Shield

The analog signal processing was realized as a layer (called "shield" in context with ARDUINO©) to be compatible with the ARDUINO©DUE, a micro-controller board based on the Atmel SAM3X8E ARM Cortex-M3 central processing unit (CPU) with 12 analog input pins, each of which can provide 12 bits of amplitude resolution. The layer design enables the system to be stacked and therefor to record up to 6 bio-signal channels.
The system was designed in CadSoft EAGLE™PCB Design Software and simulated in the SPICE-based analog simulation program TINA-TI by Texas Instruments©. The design of the active filter was done using Texas Instruments' software Filter Pro™.

The main parts in this electronic circuit are the amplification and filter stage, the half-wave rectification stage and the power supply circuit. The complete circuit layout and the printed circuit board (PCB) are shown in A.2 and A.1. Pictures 3.2 and 3.3 show the bare, drilled PCB

and the assembled PCB shield already stacked to the micro-controller board. The PCB has a double-layer design and its dimensions are 56 mm × 83 mm. All used components except the dual in-line package (DIP) switches to select the channels are surface-mount devices (SMDs). The active components have a small outline integrated circuit (SOIC) package size and all passive components, like resistors, capacitors and inductors have package sizes of either 1206 (3.2 mm × 1.6 mm) or 0805 (2.0 mm × 1.2 mm). The USB plug to connect the electrode strap is a type-B Mini-USB connector. To allow stackability extra long shield stacking headers are used.

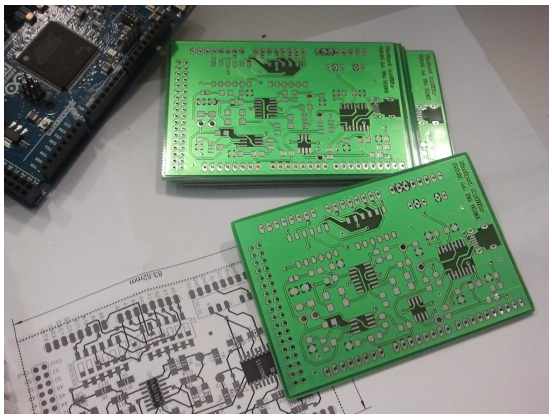


Figure 3.2: Bare, drilled PCB of the bio-signal shield.

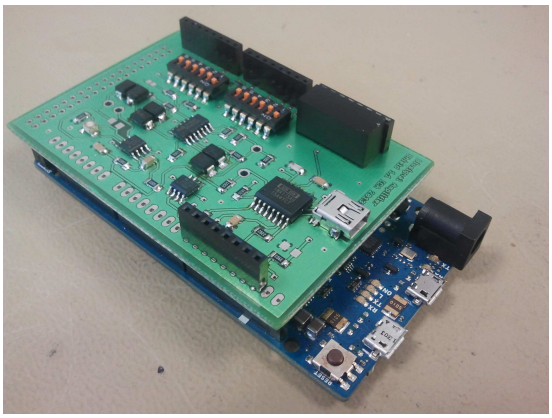


Figure 3.3: Assembled PCB of the bio-signal shield already connected to the ARDUINO© DUE micro-controller board.

Amplification and Filtering

The circuit of the amplification stage is shown in figure 3.4. The electrode inputs E1 and E2 are connected to the inputs 4 and 5 of the instrumentation amplifier, the integrated circuit (IC) 1, an INA115B by BURR-BROW©. This chip has very low offset voltage, a very high CMRR and a wide supply voltage range of ±2.25V to ±18V. This makes this chip perfect for the use in battery-operated devices [6]. The instrumentation amplifier can be divided into two stages, the buffer and the differential amplification stage.
The two operational amplifiers and the two resistors R on the left side inside the IC1 operate as the buffer stage. This stage provides a high input impedance, increases the CMRR of the circuit and enables the buffers to handle much larger common-mode signals without clipping. Furthermore, the IC INA115 has an over voltage protection included on the input side. The two output functions of the buffer stage are shown in equations 3.1 and 3.2.

$$V_1 = V_4(1 + \frac{2R}{R_G}) \tag{3.1}$$

$$V_8 = V_5(1 + \frac{2R}{R_G}) \tag{3.2}$$

V_1 and V_8 are the electrode input signals, V_4 and V_5 are the output voltages of the buffer stage amplifiers, R are internal resistors with a fixed value of 25 kΩ and R_G is the gain resistor, which connection pins are led outside the chip for gain adjustment. Equation 3.3 shows the relation between the gain of the buffer stage G_{ina} and the selected resistor R_G. As the CMRR of the

system increases with higher gain, R_G is chosen to be 422 Ω, which results in a gain factor of 120 ($\sim$ 41.5 dB) and a CMRR of $\sim$ 120 dB.

$$G_{ina} = 1 + \frac{2R}{R_G} \tag{3.3}$$

The third operational amplifier and four additional resistors R with the fixed value of 25 $k\Omega$ form the differential amplification stage inside IC1. A differential amplifier can be seen as a combination of an inverting and a non-inverting amplifier. Under the precondition that all resistors have the same value and therefor the gain of this stage equals 1 its output function results in a simple subtraction of the two input signals V_{in2} at pin 8 and V_{in1} at pin 1 of the differential amplifier. Equation 3.4 shows the output function of the single differential amplifier.

$$V_{dif} = V_{in2} - V_{in1} \tag{3.4}$$

As all internal resistors of the INA115 have the same value, the combination of the buffer stage and the differential stage results in the transfer function for the instrumentation amplifier that is shown in equation 3.5.

$$V_{out} = V_{11} = (V_4 - V_5)(1 + \frac{2R}{R_G}) \tag{3.5}$$

Connecting a capacitor C_1 to R_G against ground results in a passive RC low-pass filter. This is done to suppress noise beyond the useful frequency range of EMG signals. Equation 3.6 shows the calculation of the capacitor value and the relation to the cut-off frequency f_{c_1} of the filter. Choosing C_1 to be 330 pF yields $f_{c_1} = 1.2$ kHz.

$$f_{c1} = \frac{1}{2\pi\tau} = \frac{1}{2\pi R_G C_1} \tag{3.6}$$

A so called DRL circuit is realized using IC2, a 5 V double-supply optimized, micro-power IC OPA2241 by BURR-BROWN©. This name is historically evolved as the reference electrode for electrocardiography (ECG) was commonly placed at the patient's right leg. ECG is a technique similar to EMG, but for measuring heart pulses. This DRL circuit is inverting the error signal and feeding it back to the users body. By choosing the two resistors R_4 and R_5 to have the same value (here 390 kΩ), the gain of this inverter $G_I = 1$.

The output signal of the circuit shown in figure 3.4 is now again amplified and filtered by the operational amplification circuit shown in figure 3.5.
The capacitor C_6 eliminates all direct current (DC) offset caused by varying body ground potentials. Measurements made on a circuit built up on an evaluation board have shown that this offset can vary between 0 V and ±10 μV.
The signal is then fed to an active, second order, multiple feedback, Butterworth low-pass filter with a cut-off frequency $f_{c_2} = 1$ kHz and a gain of 6 ($\sim$ 15,5 dB). The operational amplifier and filter circuit is inverting the signal and has a band-pass ripple of 1 dB and a stop-band attenuation of 12 dB per octave. The transfer function and the calculation of the cut-off frequency f_{c_2} of this filter is shown in equations 3.7 and 3.8 [7].

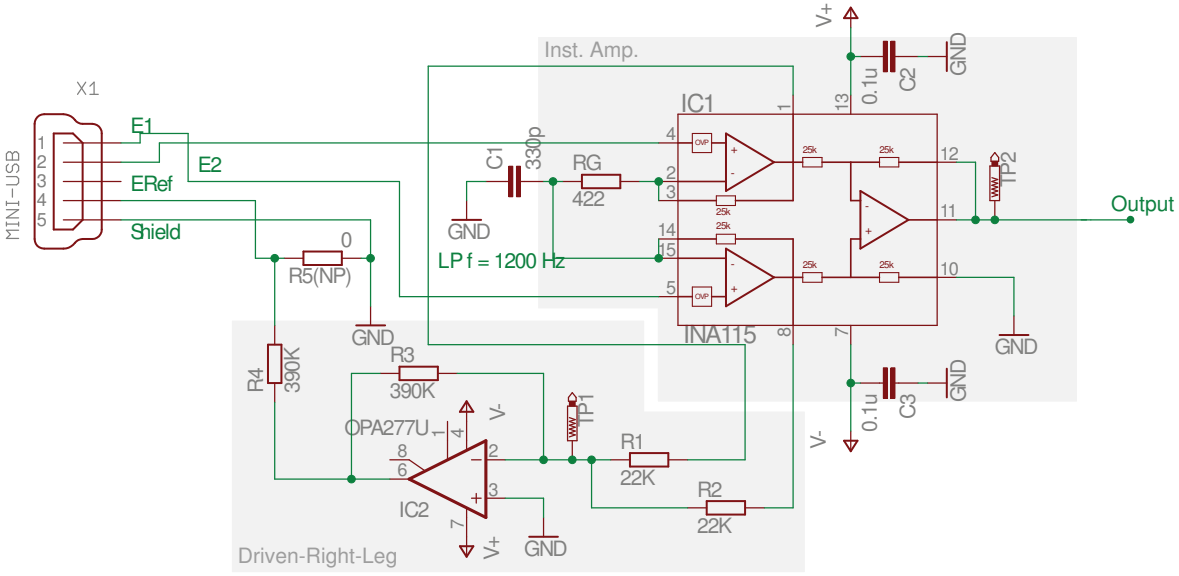


Figure 3.4: Instrumentation amplification circuit + filtering

$$\frac{V_{out}}{V_{in}} = \frac{-\frac{1}{C_4C_5R_6R_7}}{s^2 + s\frac{1}{C_4}(\frac{1}{R_6} + \frac{1}{R_7} + \frac{1}{R_8}) + \frac{1}{C_4C_5R_7R_8}} \tag{3.7}$$

$$f_{c2} = \frac{1}{2\pi\sqrt{R_7R_8C_4C_5}} \tag{3.8}$$

In the next step a passive limiter, consisting of two Schottky-diodes, is terminating the amplified signal to ±0.3 V. On the one hand, this is done to avoid unwanted high voltage signals (e.g. peaks) to get amplified in the next stage and may harm electronic circuits and devices that are connected to the output of the bio-signal shield. On the other hand, it is done to not exceed the consumer audio line level of −10 dBV and a peak amplitude of 0.447 V. This assures the compatibility with commercially available audio interfaces as the amplified and filtered signal at test point (TP) 4 can be connected to their input. EMG levels measured during activation of the sternocleidomastoid muscle are ranging from 0 V to 90 μV (V_{max}) [1]. This muscle is one of the biggest muscles in the front side of the neck. Therefor no higher levels are expected measuring sEMG muscle activation during speech. Multiplying the gain factor of the instrumentation amplifier $G_{ina} = 120$ and of the active low-pass filter $G_{filt} = 6$, the amplification factor of the whole system till TP 3 can be determined: $G_{TP3} = G_{ina}G_{filt} = 720$. So maximum EMG levels L_{TP3max} are expected:

$$L_{TP3max} = V_{max}G_{TP3} = 10,8mV \tag{3.9}$$

For the measurement and recording of bigger skeletal muscles, e.g. thigh or forearm muscles, expected sEMG levels are ranging from 0 V up to 1 mV. At an input level of 1 mV the amplified signal (0.72 V) at TP 3 is already clipping and will be attenuated to a level of 0.3 by the

limiter. So EMG signals up to a level of 416 μV are amplified correctly and not limited. For the measurement of bigger skeletal muscles either the gain resistor of the instrumentation amplifier R_G or the gain factor of the active low-pass filter has to be adapted.

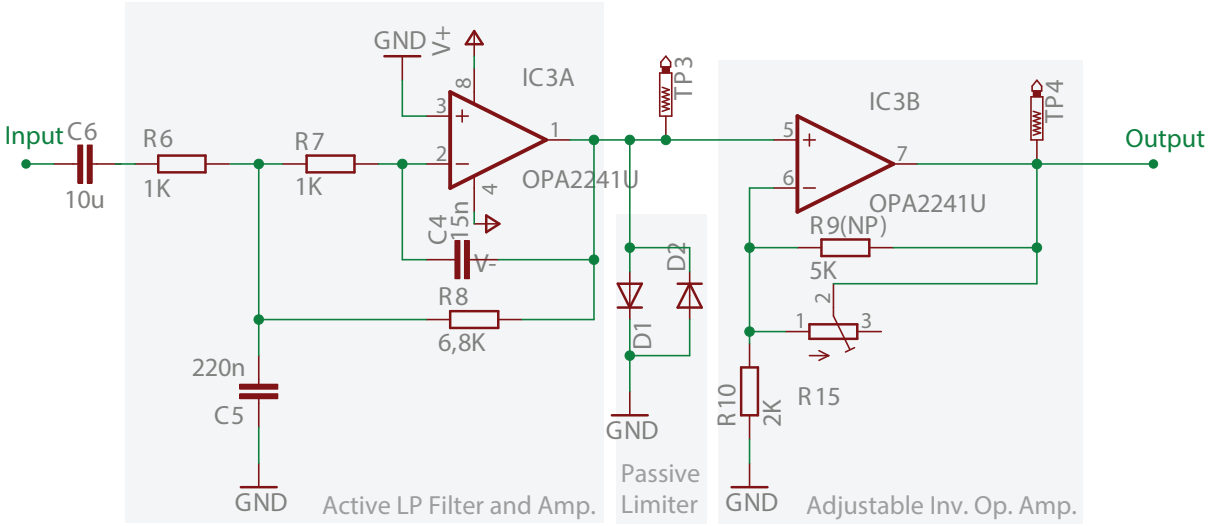


Figure 3.5: Operational amplification circuit + filtering.

Based on the actual gain configuration for sEMG measurement during speech, the limiter is followed by an inverting operational amplifier. On its output, the signal is again inverted and amplified to compensate the signal inversion of the active low-pass filter. The amplification factor can be adjusted manually in the range of 1 to 5. That yields an overall system amplification factor at TP 4 between 720 ($\sim$ 57.1 dB) and 3600 ($\sim$ 71.1 dB).

Half-wave Rectification

In the half-wave rectification stage, the signal is split into its positive and negative part. The negative part is inverted to become positive and both signals are sent to the analog inputs of the micro-controller. This stage was implemented to double the amplitude resolution of the digitized signal. The analog-digital converters (ADCs) of each input can convert the signal amplitude in 12 bit resolution (4096 steps), which in literature is mentioned to be the minimum recommended resolution [14]. Therefor, the positive and negative half-wave are converted separately and put together later in digital domain running a simple line of code on the ARDUINO©board. This method allows the analog-digital conversion of the whole signal to be performed in a higher amplitude resolution of 13 bit (8192 steps). The rectification circuit is shown in figure 3.6.
The rectification circuit is built out of three operational amplifiers of the type OPA4241 by BURR-BROWN©, a low power dual-supply voltage IC in quad version (four operational amplifiers are included in one IC package with 14 pins). IC4A and IC4C are used to form the two precision half-wave rectifiers, each including two Schottky-diodes (D_3, D_4 and D_5, D_6) and two resistors (R_{13}, R_{14} and R_{11}, R_{12}). The operational amplifiers in this circuit are wired to be inverting operational amplifiers. Depending on the orientation of the diodes the rectifiers either let the positive (IC4A) or the negative (IC4C) half-wave pass and eliminates the negative (IC4A) and the positive (IC4C) half-wave, respectively. For example, in the circuit at IC4A,

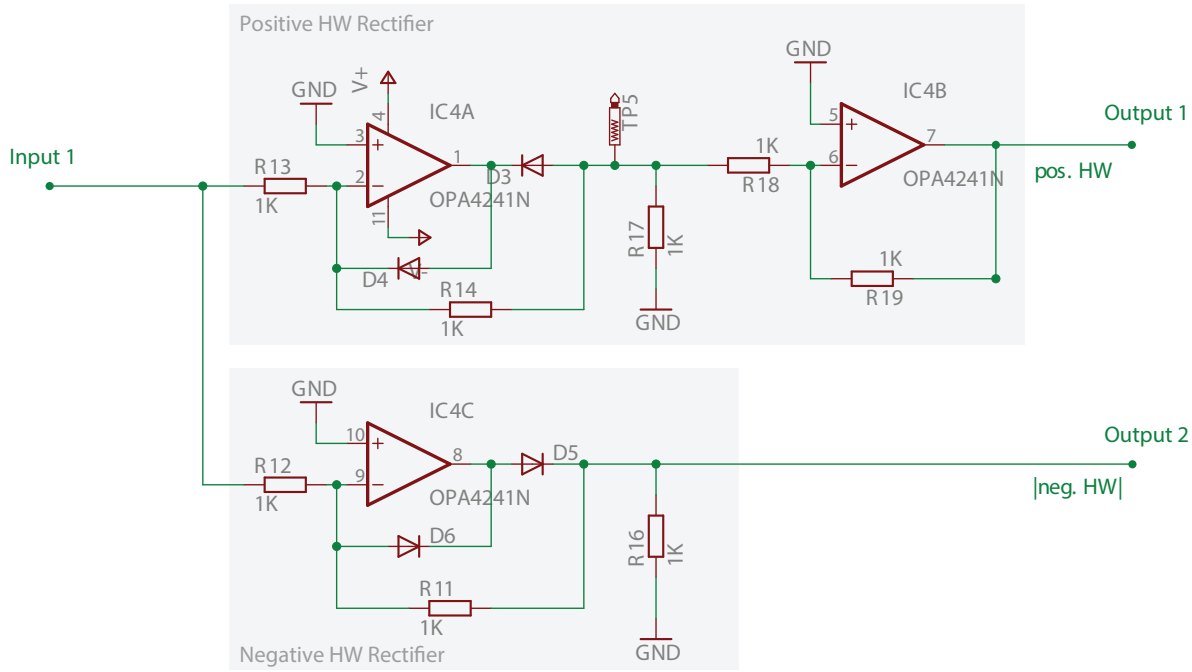


Figure 3.6: Combination of two half-wave rectification circuits.

the positive half-wave is kept and the negative half-wave is eliminated, if the input is greater than zero. In this case, the diode D_3 is on and D_4 is off, so the circuit works as a common inverting amplifier with the amplification factor of $-\frac{R_{14}}{R_{13}}$. If the input is less than zero, D_3 is off and D_4 is on. Then there is no current through R_{14} and so the negative half-wave is suppressed. The negative half-wave rectifier works the same way, but with reciprocally orientated diodes. As it is required to have both signals to be positive, the inverted positive half-wave (TP5) has to be inverted again. This is done using an inverter, which is realized using a simple inverting operational amplifier with the amplification factor of $-\frac{R_{19}}{R_{18}} = -1$. Here no latency is expected as the slew rate of the operational amplifier is very low and is not effecting signals in the frequency range of EMG measurement. Finally the positive half-wave is present on output 1 and the inverted negative half-wave is present on output 2.
Connecting the bio-signal shield to the micro-controller board, these two outputs are ready for analog-digital conversion.

Power Supply

As all the active parts used in the circuit are designed for dual power supply in the range of ±2.5 V to ±18 V the power supply of the micro-controller board (5 V) is used to power the shield. To transform the 5 V into ±5 V a 2 W DC/DC converter by Traco Power© is used.

Simulation

As mentioned in section 3.1 the circuit of the bio-signal shield was simulated using the SPICE-based analog simulation program TINA-TI by Texas Instruments. A screenshot in figure 3.7 shows the simplified simulation of the circuit. Some active components are not available in the simulation software, hence these elements are replaced by elements with similar characteristics. For example, the instrumentation amplifier is replaced by an IC INA118, which only differs from the IC INA115 in the used values of the internal resistors and the rectification amplifiers are

replaced by ICs TL082, which are very common and often used standard operational amplifiers. Test signals in the simulation are sine waves of 30 μV (peak) and 40 μV (peak) with a frequency of 100 Hz and 170 Hz. The oscilloscope screen in figure 3.8 shows the running simulation with the two tapped signals at VF8 and VF9 which corresponds to the signals at output 2 and TP 5 in the realized circuit layout shown in figure 3.6. It can be seen that the subtraction in the instrumentation amplifier stage as well as the amplification in the mark-up stage (U2) works fine. The splitting and therefor the rectification of the signal works satisfying as well.

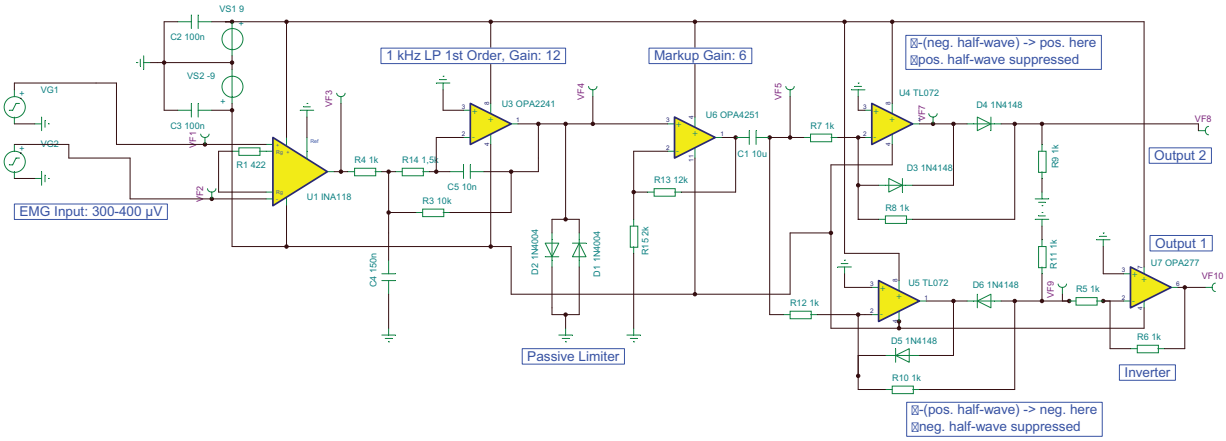


Figure 3.7: Simplified simulation setup of the hardware circuit.

The blue line in the simulation screen represents VF8, the inverted negative half-wave (values greater then zero) and the green line represents the inverted positive half-wave (values less then zero). Plotting both signals into one screen results in the visual summation of both signals. The output is a split signal of the inversion (phase shift of 180 degrees) of the amplified signal VF4 in the simulation corresponding to the signal at TP 4 in the realized circuit.

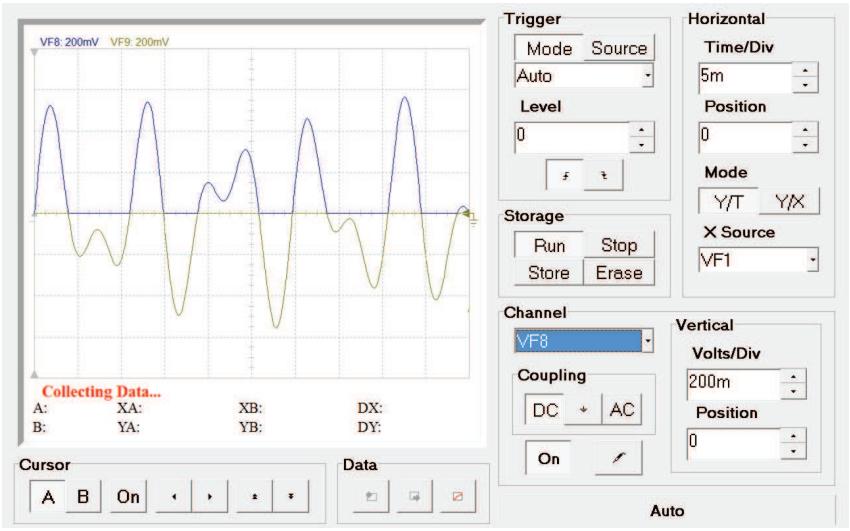


Figure 3.8: Oscilloscope screen of the simulation, showing two signals at TPs VF8 and VF9.

Version History

The first and realized version of the bio-signal shield is version 1.1, presented in pictures 3.2 and 3.3. Due to a lack of one connection between the resistors $R12$ and $R13$ and some smaller

possible improvements this design was revised. A detailed list of the change-log is presented in list A.1.2. The new version 1.2 is the base for all explanations in this work, including circuits and the PCB layout.

3.2 Recorded Database

The recorded database consists of 114 German sentences. Recordings were made with one female and one male speaker under three scenarios: sEMG signal and EL speech signal, sEMG signal and healthy speech signal and sEMG signal while silent speaking.
During the recordings, the test subject was overseen by a supervisor. The supervisor had the control over the speech recording software in order to control and modify the recording process immediately. The used software was *SpeechRecorder* [9] which had been designed to record speech corpora and databases. The test subjects were recorded sitting in a meeting room with a low reverberation time. The test subject had to speak sentences displayed on a screen. EMG and speech signals are recorded using either the bio-signal shield connected to an audio interface (RME Fireface 800) at TP4 (the filtered and amplified but not rectified signal) and a head-mounted microphone AKG©HC 577 L with omni-directional pickup pattern. The microphone was chosen to ensure a consistent recording quality, since it guarantees a constant distance of about 2 cm from the corner of the mouth. The sentences are recorded in stereo where the left channel contains the EMG signal and the right channel contains the speech signal. The audio interface ensures a very good and high standard digital signal quality. The sampling rate of the audio interface was set to 48 kHz and the bit rate was set to 24 bit. Figure 3.9 shows the block diagram of the recording setup.

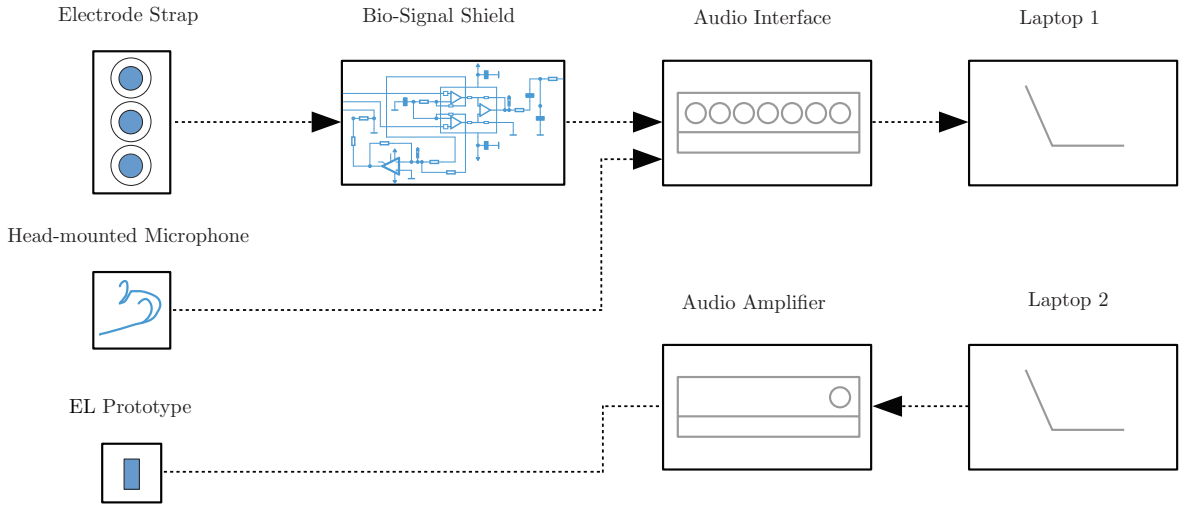


Figure 3.9: Block diagram of the EMG and speech database recording setup.

The amplified sEMG signal as well as the microphone signal are recorded by the audio interface which is connected to laptop 1. On laptop 1 the software *SpeechRecorder* manages the recording and the visual presentation of the sentences. Laptop 2 is playing the excitation signal, which is amplified by an hi-fi audio amplifier and fed to the EL prototype mounted at the subject's neck. A picture of the recording setup is shown in figure 3.10.
As the focus in the design of the bio-signal shield was the development of a low-cost, battery-operated real-time system with a standard sampling rate of $\sim$ 8 kHz and a bit rate of 13 bit, the recording with the professional audio interface for better signal quality was preferred. The EL speech was recorded using a prototype EL held against the neck. All recordings with EL speech are made by healthy speakers simulating the disability to speak by holding their breath. sEMG

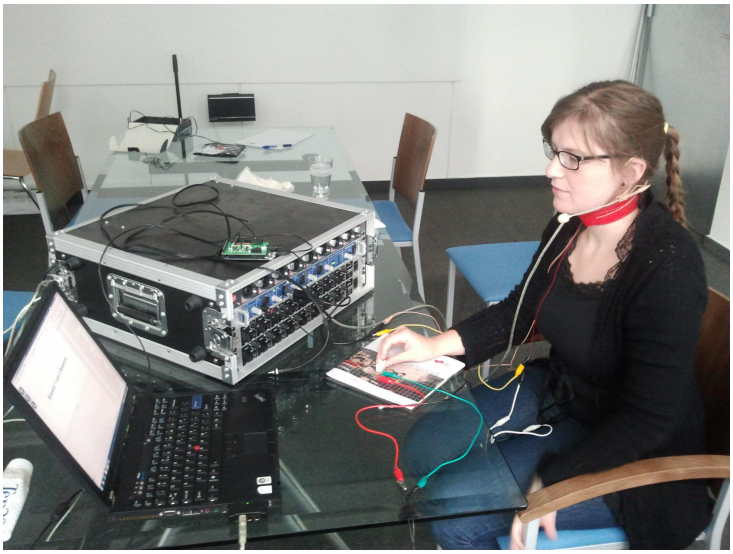


Figure 3.10: sEMG and speech database recording setup.

signals that occur from silent articulation are recorded to expand the database of healthy speech for further investigations into sEMG signals of disordered speech after total laryngectomy. The speech database consists of phonetically balanced sentences and is split into a development and an evaluation set. The development set consists of nine sentences and is used for parameter tuning and testing scenarios. The tested settings are applied on the sentences of the evaluation set. All data plots in section 5.2 are based on analyzing the evaluation set. In tables A.1 and A.2 all sentences of the two data sets of the database are presented.

4 Proposed Method

4.1 Introduction

In addition to the documentation of the hardware development and the signal acquisition in chapter 3, this chapter describes the second main area of this work, the processing of the digitized sEMG signal and the evaluation of different EMG activity detection methods with the aim to enhance EL systems by automating the on/off switch of the device. Figure 4.1 shows a data flow diagram of the whole processing, detection and evaluation system.
In a first step, different preprocessing methods are applied, including adaptive noise cancellation (ANC) to decrease the interferences caused by the excitation signal of the EL held against the neck in close distance to the sEMG electrodes. A 50 Hz notch filter reduces possible crosstalk by power line hum and a filter-bank splits the clean signal into several filter bands for further processing. These operations as well as the downsampling of the signal are presented in section 4.2. The next section 4.3 is dealing with the preparation of the signal for a following activity detection (AD). In this part three different envelope calculation methods (ECMs) (root mean square (RMS), Hilbert transform (HIL) and Teager energy operator (TEO)) in combination with several different smoothing window lengths are applied on the signal. The classification into active and non-active parts is done using the three AD algorithms single threshold detection (STD), double threshold detection (DTD) and adaptive threshold detection (ATD). These algorithms are presented in section 4.5. To compare the determined detection results to "true" values, the corresponding speech signals are analyzed and ground truth (GT) files, including the manually defined on- and offset times, are generated. In the error calculation stage the on/off results are compared to these GT values. Four errors, dependent on their appearance in relation to the on- and offset times of the GT and one not time-dependent error are calculated (see section 4.6).

The processing, detection and evaluation system presented in this section is implemented in MathWorks MATLAB©except for the manually definition of the GT times which was done using the audio editing software AUDACITY©.

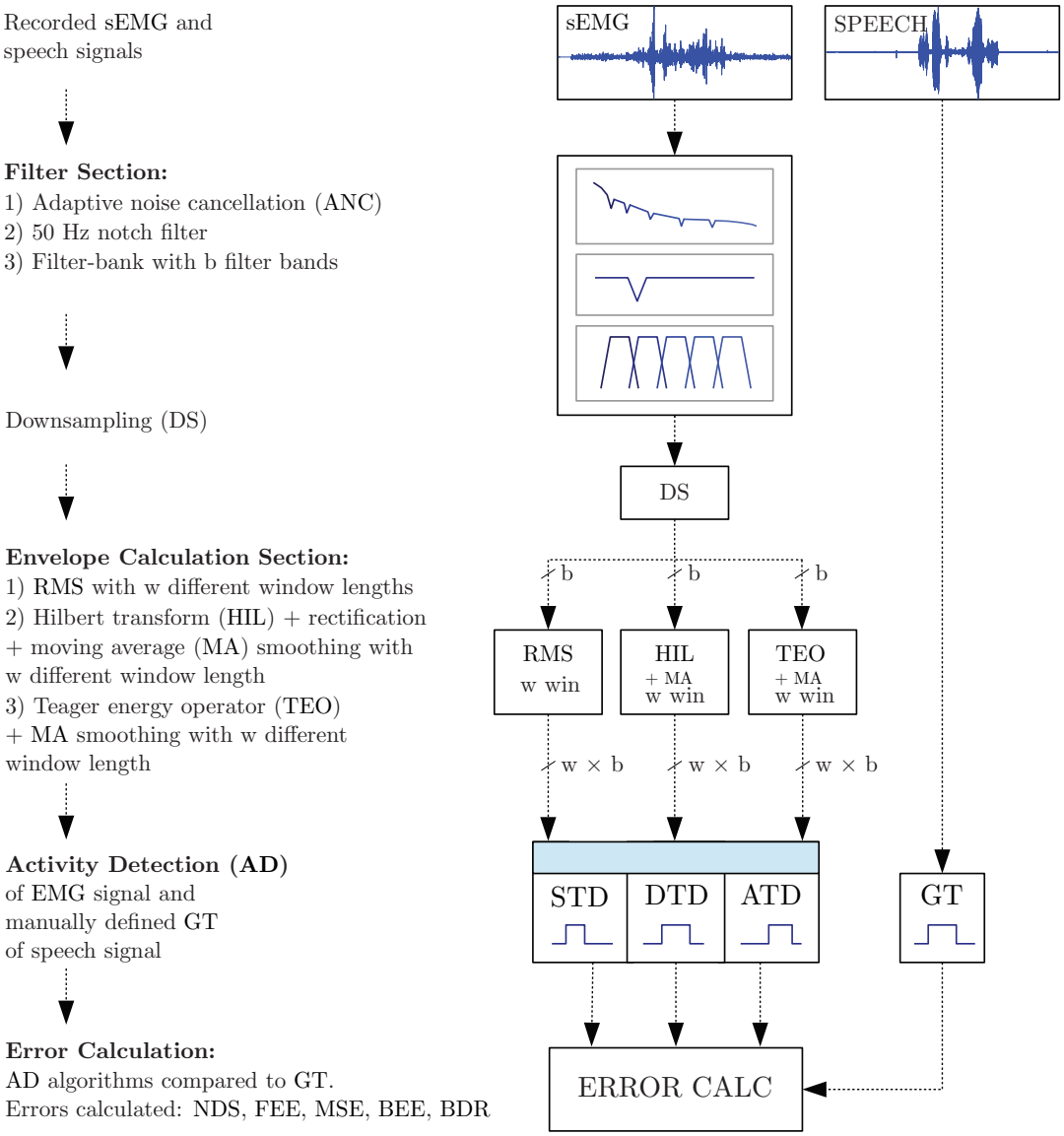


Figure 4.1: Flow chart of the EMG processing, detection and error calculation system

4.2 Preprocessing

The preprocessing of the signal consists of three stages to clean and prepare the recorded sEMG signal for subsequent processing and classification steps.

4.2.1 Adaptive Noise Cancellation

The recording of sEMG signals at the neck while speaking with the hand-held as well as the prototype EL yields a crosstalk of the excitation signal and the sEMG signal. Thus, the excitation signal is also recorded by the sEMG electrodes, although with a small amplitude. To prepare the signal as good as possible, this crosstalk has to be reduced to a minimum.
The excitation signal used in the recordings is generated following the Liljencrants–Fant (LF) model [10], which is a common pulse model used in voice production. It is a waveform model where the flow derivative is described parametrically. The shape of the LF model is defined by four independent parameters T_p, T_a, T_e and E_e. The derivative can be divided in three phases of the glottal cycle: The opening phase from T_0 – T_p ($T_0 = 1/F_o$), the return phase T_p – T_e

and the closed phase $T_c - T_0$ [15]. The mathematical definition of the LF model is presented in equation 4.1. The waveform and its derivative is shown in figure 4.2.

$$V(t) = \begin{cases} E_0 e^{\alpha t} sin\omega_0 t & \text{for } t < T_e \\ \frac{-E_0}{\varepsilon T_a} . \left[e^{-\varepsilon(t-T_e)} - e^{\varepsilon(T_c-T_e)} \right] & \text{for } T_e < t < T_c \\ 0 & \text{for } T_c < t < T_0 \end{cases} \tag{4.1}$$

with

$$\omega_0 = 2\pi f_0 = 2\pi \frac{1}{T_0} \tag{4.2}$$

where $U_g(t)$ represents the waveform amplitude, U_0 represents the maximum amplitude and T_d represents the slope of the closing phase of the waveform. $-E_e$ represents the negative maximum of the velocity $V_g(t)$.

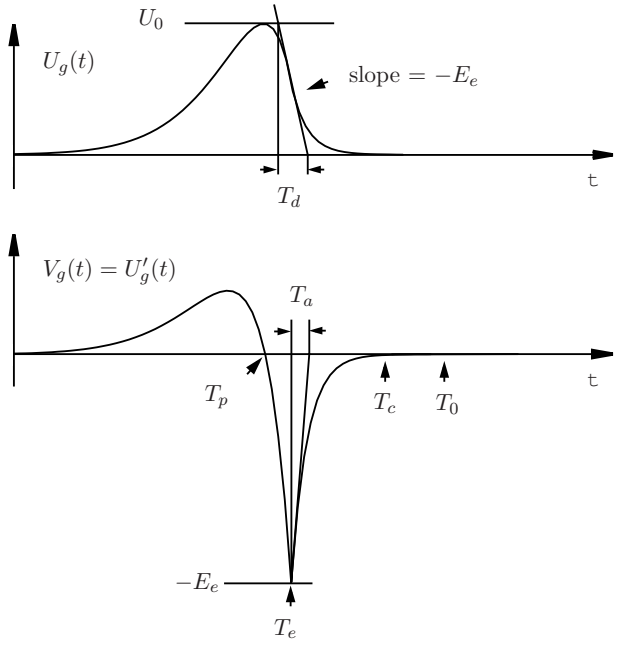
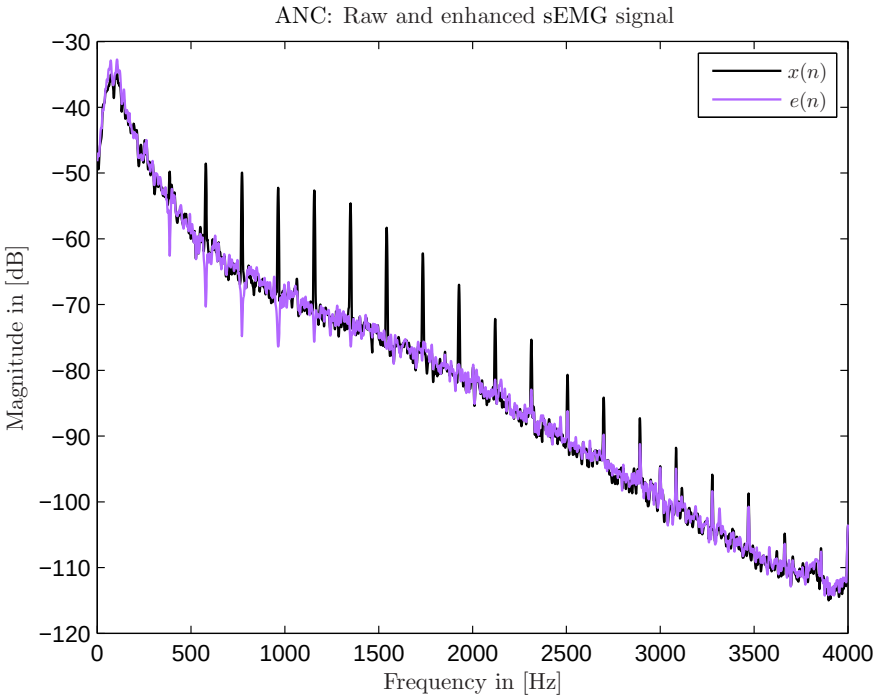


Figure 4.2: Glottal flow $U_g(t)$ and its derivative $V_g(t)$.

As the excitation signal is known, it can serve as a reference input signal for ANC. This method was used to achieve the desired reduction of the crosstalk. The excitation signal $n_1(n)$ and the recorded crosstalk $n_2(n)$ in the sEMG signal are correlated and the sEMG signal $s(n)$ is uncorrelated with both of the signals. Thus a noise cancellation filter with the input of $n_1(n)$ is rebuilding a signal that is as close as possible to $n_2(n)$. This signal is subtracted from the delayed input signal $x(n - N_c) = s(n - N_c) + n_1(n - N_c)$ which results in a decrease of the crosstalk signal in the desired sEMG signal $e(n)$. Two parameters which have to be tuned depending on the characteristics of $n_1(n)$ are the signal delay N_c and the filter length L [30].

This method is successfully used for improving headphone systems and for the enhancement of ECG recording of unborn children to reduce the influences by the mothers heart pulse [12]. Figure 4.3 shows the frequency spectrum of the raw sEMG signal $x(n)$ and the enhanced signal

$e(n)$. The black peaks represent the excitation signal's fundamental frequency and its harmonics. These are eliminated by the ANC algorithm. It can be seen that the attenuation of the peaks works well below 2.5 kHz. Above this frequency the signal is not relevant in terms of sEMG measurement, as these high frequencies have magnitudes lower than -70 dB and no information is expected there. Therefor, the sample rate is reduced from 48 kHz to 16 kHz.



:

Figure 4.3: Frequency spectrum of the raw and noise canceled signal.

4.2.2 Notch Filter

To reduce interferences from electric hum (main hum) caused by magnetic fields close to the sensor and amplification unit, a notch filter with $f_c = 50$ Hz is applied on the signal. This filter has a narrow bandwidth of 5 Hz and thus is not significantly reducing the signal energy but the unwanted noise in this frequency band. In earlier development board-built versions of the amplification and filter circuit presented in section 3.1.2, the notch filter for the digitized signal was very important. Due to the improved ground management and shielding in the newest version 1.2 of the bio-signal shield, the interference from electric hum are extremely low and the importance of the notch filter declined. If higher interference from electric hum occurs the notch filter could be replaced by adaptive line enhancement (ALE), an adaptive system where compared to ANC no reference signal is needed and where the EMG signal energy is hardly affected [8].

4.2.3 Filter-bank

The used filter-bank is a set of 2 low-pass (filter 1 and 2) and 5 band pass filters (filter 3 – 7), all Butterworth filters of the order $N = 2$. The full range signal (called filter band 1) is split into 6 filter-bands for comparison of the influence of different frequency ranges in the subsequent used algorithms. The transfer functions of the filter bank are presented in 4.4 and its frequency ranges are listed in table 4.1.

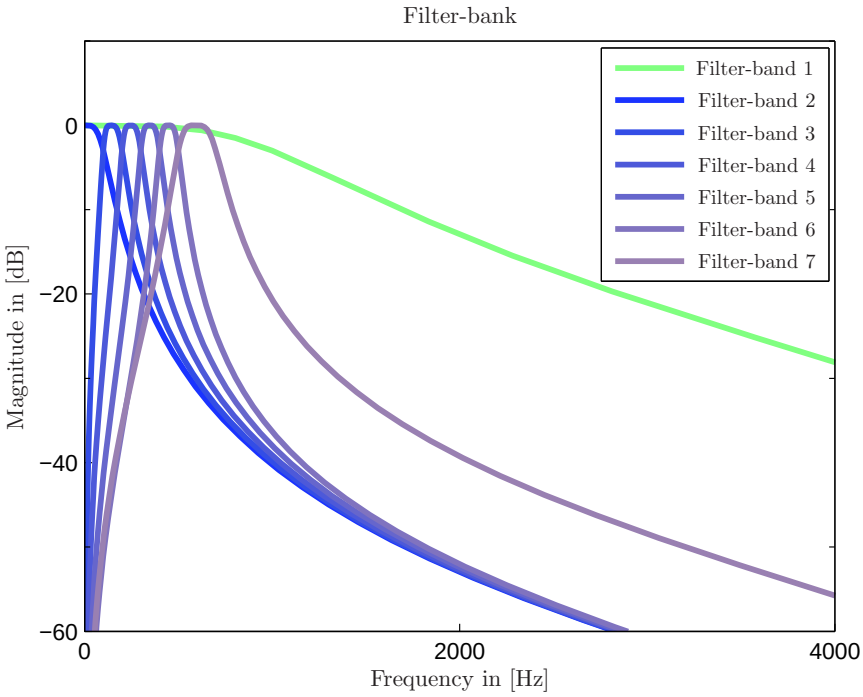


Figure 4.4: The signal is split into 6 filter-bands (filter-band 2 – 7). The transfer function of the full range (green line) is called filter-band 1.

Table 4.1: Filter-bank frequency ranges

Band	Filter Type	Frequency Range
Filter-band 1	low-pass	0 Hz – 1000 Hz
Filter-band 2	low-pass	0 Hz – 100 Hz
Filter-band 3	band-pass	100 Hz – 200 Hz
Filter-band 4	band-pass	200 Hz – 300 Hz
Filter-band 5	band-pass	300 Hz – 400 Hz
Filter-band 6	band-pass	400 Hz – 500 Hz
Filter-band 7	band-pass	500 Hz – 700 Hz

4.3 Envelope Calculation

The enhanced sub-band signal is the input signal for the feature extraction section using three ECMs. The extracted features are the windowed RMS, the HIL and the TEO of the signal.

4.3.1 Root Mean Square

In EMG analysis, the RMS of the signal is a common ECM to get the power of the signal. It has a clear physical meaning. The RMS value for a signal window of x is defined as

$$RMS\{x\} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x[n]^2} \tag{4.3}$$

where N is the number of sample points of x. By calculating a sequence of RMS values of signal blocks of the length of one defined window and an overlap, the input signal is rectified and smoothed. The number of data samples is reduced. The new sampling frequency f_{sn} is dependent on the window length win and the overlap size ol.

$$f_{sn} = f_s(win - ol) \tag{4.4}$$

4.3.2 Hilbert Transform

Another method to get the power of the EMG signal is to use the hilbert transformed signal and rectify it. The HIL amplitude is given by

$$\mathrm{HIL}\{x\} = |H\{x(t)\}| \tag{4.5}$$

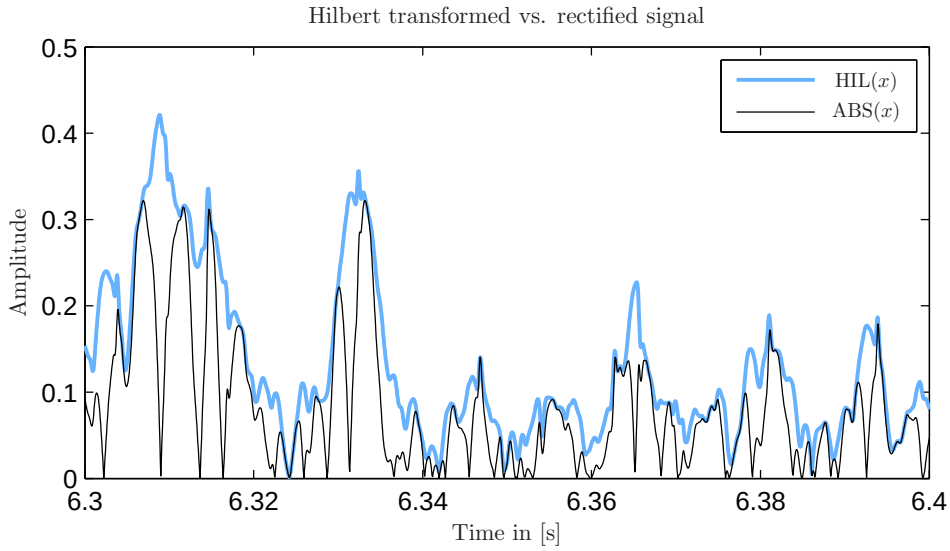


Figure 4.5: HIL envelope compared to the simply rectified sEMG signal

The HIL envelope compared to a simply rectified sEMG signal is shown in figure 4.5. It can be seen that this method works as an amplitude follower and provides, even un-smoothed, an envelope which is not touching the zero line on the y-axis.

4.3.3 Teager Energy Operator

The TEO is an operator for the instantaneous energy of a signal. For a time-discrete signal the TEO Ψ [18, 20] is defined as

$$\Psi[x_n] = x_n^2 - x_{n-1}x_{n+1} \tag{4.6}$$

4.4 Envelope Smoothing

The HIL and TEO envelopes are smoothed using a moving average (MA) finite impulse response (FIR) filter with four different window lengths win_s of 512, 1024, 1532 and 2048 samples and a hop size hs_s of 1 sample. At a sampling frequency of 16 kHz, the chosen window lengths

correspond to time values t_m of 32, 64, 96 and 128 ms. The latency τ of a step function due to MA filtering can be calculated as follows

$$\tau = \frac{t_m - 1/f_s}{2} \approx \frac{t_m}{2} \tag{4.7}$$

With $1/f_s$ being small compared to t_m, τ can be approximated as $t_m/2$ [25]. The latency also depends on the used detection algorithm, but it is important to consider the expected latency in live application for further processing or to generally try to minimize latencies. The RMS envelope is smoothed using a window $win_{rms} = win_s/2$. The overlap ol_{rms} is set to $win_s/4$. If buffering the signal in blocks of the size of the used window, a general latency of the length of the window is expected in live applications.

Applying the smoothing methods with four filter windows win_s and win_{rms}, respectively, to all seven sub-bands and signal envelopes yields a set of $4 * 7 * 3 = 84$ different features. Figures 4.6, 4.7 and 4.8 show 12 of these features in time domain. Figure 4.6 represents a full-range (filter band 1), RMS rectified sEMG signal smoothed with four different window lengths. Figures 4.7 and 4.8 present the HIL envelope as well as the TEO envelope of the signal. It can be seen that the filtering parameters win_{rms} and ol_{rms} of RMS are chosen to achieve a filter smoothness complying to the smoothness of the HIL and TEO envelopes.

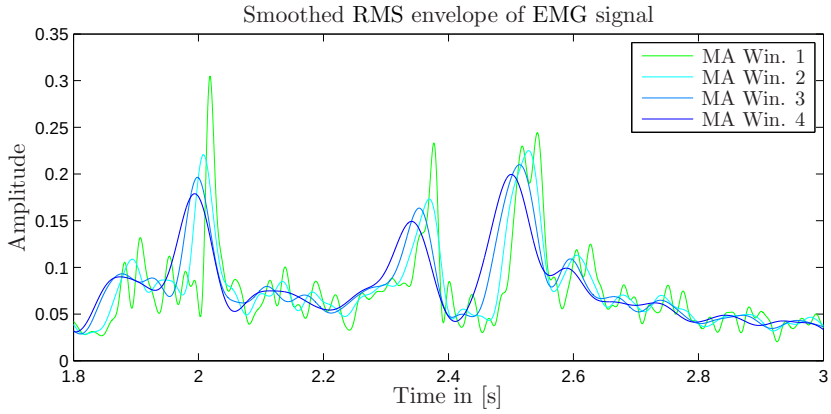


Figure 4.6: RMS of EMG signal during EL speech smoothed with four different MA windows.

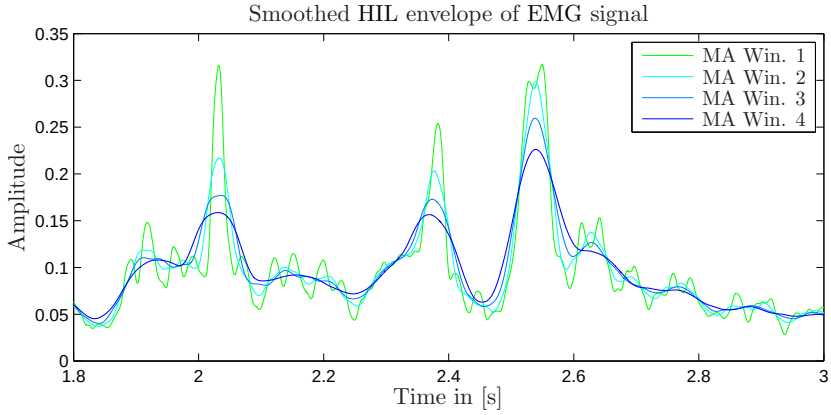


Figure 4.7: Hilbert envelope of EMG signal during EL speech smoothed with four different MA window.

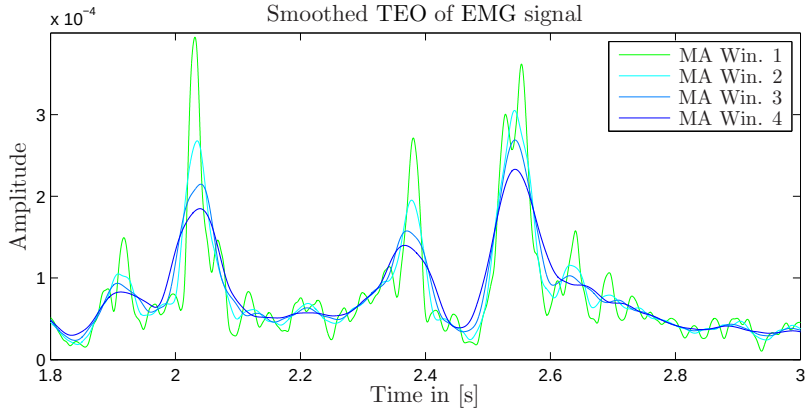


Figure 4.8: TEO of EMG signal during EL speech smoothed with four different MA windows.

4.5 Classification

For the classification of the sEMG activity of the recorded data, three activity detection algorithms are tested and evaluated. In section 4.5.1 the simplest possible detector, a STD algorithm is presented. A single predefined value, a certain percentage of the maximum amplitude, is acting as a threshold. Amplitude values above this threshold are indicating activity. The DTD presented in 4.5.2 is an improved version of the STD. By adding a second threshold, two thresholds, one for detecting the onset and the other for detecting the offset of the activity, the algorithm is less prone to small amplitude variations that occur du to the variation of the muscles firing rate. In section 4.5.3 an ATD algorithm is introduced.

4.5.1 Single Threshold Detection

The STD is the simplest possible but often used and reliably working activity detection algorithm in EMG processing. Values above the threshold are set to 1 and indicating EMG activity and values below the threshold are set to 0, indicating that there is no activity. The threshold Th_s is set to a certain percentage of the maximum amplitude of the currently analyzed EMG EL speech data. In a live scenario this predefinition of the threshold is only possible by conducting an initialization phase prior to the use of the device. In this phase, the user could be asked to perform a maximum contraction to let the system determine an appropriate threshold. Tests showed that dependent on the user, thresholds between 15 and 25 percent of the maximum amplitude yield the best detection results. An example of the detection results of this algorithm is shown in figure 4.9.

The sum over time of the detected activity blocks results in the total activity time t_{on}:

$$t_{on} = \sum_{n=1}^{k} t_n \tag{4.8}$$

where n represents the block number and k the number of detected blocks in the analyzed sentence. In figure 4.9 n a segment of the total sentence is shown, where t_n represents the first detected block in this segment.

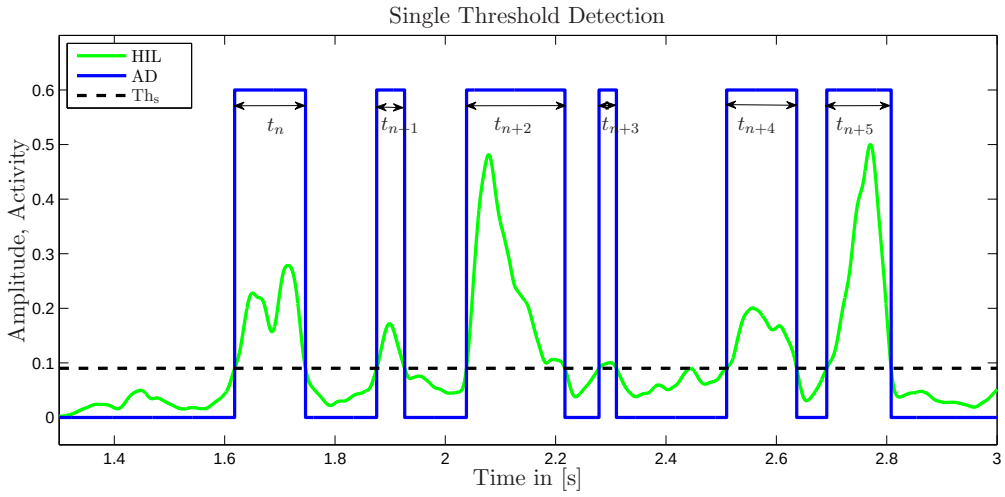


Figure 4.9: Example of the performance of the STD results, here with the threshold $\mathrm{Th_s}$ *set to 18 % of the amplitude maximum. The total detection time* t_{on} *is the sum of the lengths of all detected activity blocks* t_{n+i}.

4.5.2 Double Threshold Detection

The DTD algorithm works as a combination of two single thresholds. This algorithm is detecting activity in the sEMG signal envelope only when the amplitude is falling below a second threshold that is set to a lower amplitude level than the first one, which is detecting the onsets of activity. To avoid unwanted loss of detection due to small variations in the muscle force the lower second threshold is added. So, the DTD is an improved method of the STD and is designed to increase the reliability of the fixed STD algorithm. By making the algorithm less sensitive to small amplitude variations, smaller smoothing windows can be applied on the rectified signal. This results in smaller latency of the envelope calculation and the MA filter. The results and the thresholds of this algorithm are shown in figure 4.10.

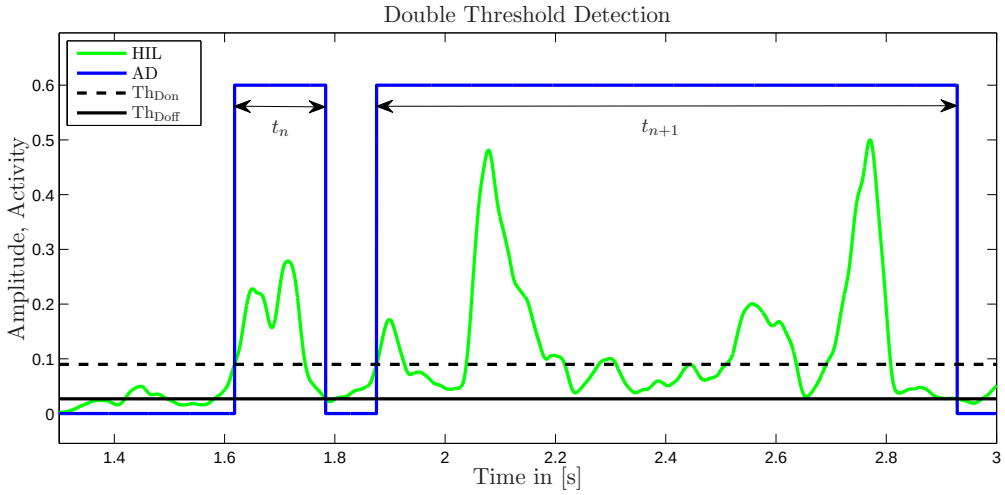


Figure 4.10: Example of the results of the DTD algorithm, here with the threshold $\mathrm{Th_{Don}}$ *set to 18 % of the amplitude maximum and the threshold* $\mathrm{Th_{Doff}}$ *set to 40% of* $\mathrm{Th_{Don}}$.

4.5.3 Adaptive Threshold Detection

To compare the two fixed threshold activity detection algorithms to an adaptive algorithm the ATD algorithm based on the *Simple Power Voice Activity Detection with Adaptive Threshold* by Petr Pollák [27] is introduced. Dependent on the power calculated in time segments of 20 ms the threshold is adapted. The maximal and minimal power $P_{max}(t)$ and $P_{min}(t)$ is calculated dependent on four parameters q_{max1}, q_{max2}, q_{min1} and q_{min2}, which define the speed of the energy update and therefor the slope of the threshold curve:

$$P_{max(t)} = \begin{cases} q_{max1}P_{max}(t-1) + (1-q_{max1})P(t) & \text{if } P(t) \geq P_{max}(t-1) \\ q_{max2}P_{max}(t-1) + (1-q_{max2})P(t) & \text{if } P(t) < P_{max}(t-1) \end{cases} \tag{4.9}$$

$$P_{min(t)} = \begin{cases} q_{min1}P_{min}(t-1) + (1-q_{min1})P(t) & \text{if } P(t) \leq P_{min}(t-1) \\ q_{min2}P_{min}(t-1) + (1-q_{min2})P(t) & \text{if } P(t) > P_{minx}(t-1) \end{cases} \tag{4.10}$$

where $P(t)$ represents the power of the t-th time segment. The algorithm is classifying active and non-active parts of the signal by comparing $P(t)$ to the calculated threshold $\mathrm{Th_A}$. This threshold is given by:

$$\mathrm{Th_A} = P_{min}(t) + \frac{p}{100}.(P_{maxt}(t) - P_{min}(t)) \tag{4.11}$$

The algorithm is tuned to decrease the threshold slowly during low power segments and to increase the threshold level fast during segments with higher power. As this algorithm should also work in live scenarios the powerful voice activity detection (VAD) post-processing proposed by Pollák is not implemented. An AD post-processing method for the use in live-scenarios is presented in section 4.5.4. The results of the algorithm are shown in figure 4.11.

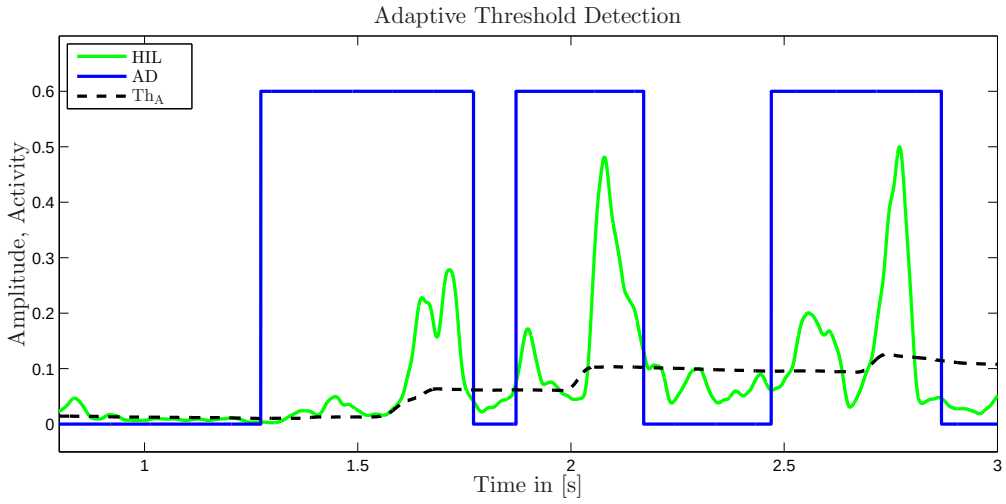


Figure 4.11: Example of the results of the ATD algorithm. The threshold $\mathrm{Th_A}$ *is adapted by analyzing the energy in signal segments of 20 ms.*

4.5.4 Time Constant Detection Smoothing

For post-processing of short-time VAD, it is common to apply detection smoothing or long time detection algorithms on the detection output. This is done to avoid short interruptions of the detection. In live scenarios this is problematic as for offset events the length of the following interruption cannot be predicted. As mentioned in section 2.3, sEMG signals during speech show a pre-activation compared to the speech signal. A proposed algorithm, the time constant detection smoothing (TCDS) algorithm is using this period of 40 ms to smooth the detection output in a real-time application and therefore avoid small interruptions of detection. In the same step the algorithm is compensating EMG pre-activation. TCDS is delaying every on- and offset by the time constant $t_{tcds} = 40ms$. The performance of this algorithm is shown in figures 4.12 and 4.13.

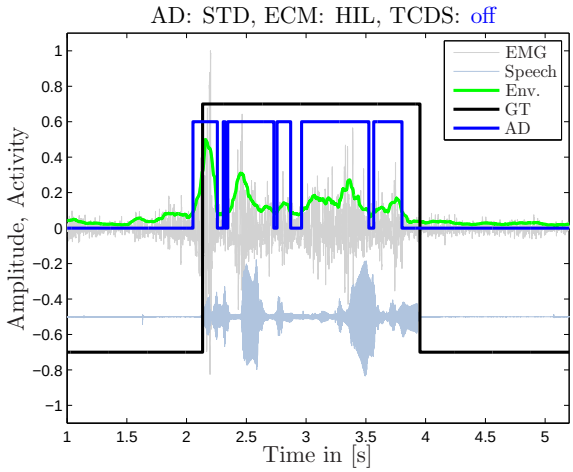


Figure 4.12: Detection output with TCDS turned off compared to GT.

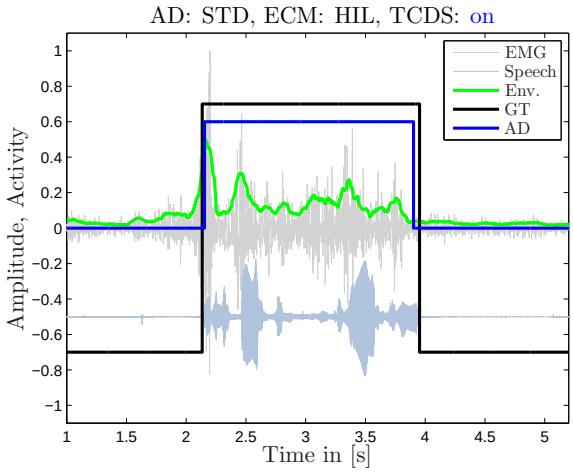


Figure 4.13: Detection output with TCDS turned on compared to GT.

A recorded sentence of a male speaker of the length of 5.3 s is analyzed using the smoothed Hilbert envelope (smoothing window 3), full-range (filter band 1) signal. The speech signal waveform, presented in blue, as well as the GT, presented in black is shifted by -0.5 for better display. The STD algorithm is classifying the envelope (green) in terms of activity. It can be seen that the detection vector is quite hackly compared to the GT vector (see figure 4.12). The real-time TCDS is improving the detection to a nearly perfect result (see figure 4.13).

4.6 Error Calculation

The on/off results of all sentences and all filter bands, filtering/smoothing methods and detection algorithms are compared to speech GT and errors regarding the correctly detected activation time and depending on the interruptions of the detection are calculated.

4.6.1 Timed Errors

The main error is separated into four different errors dependent on time appearance of the error compared to the GT. This is done to get time information about the error [5]. Different methods yield different distributions of these separated errors, although the total error is constant. The total error is classified in the following errors: front end error (FEE), back end error (BEE), middle speech error (MSE) and noise detected as speech (NDS). Equation 4.12 shows the error

classification dependent on the onset time of the GT t_{GTon}, the offset time of the GT t_{GToff} and a time corridor $t_{corr} = 200$ ms before and after the on- and offset events.

$$err(t) = \begin{cases} \text{FEE}(t) & \text{if } t_{GTon} - t_{corr} < t < t_{GTon} + t_{corr} \\ \text{BEE}(t) & \text{if } t_{GToff} - t_{corr} < t < t_{GToff} + t_{corr} \\ \text{MSE}(t) & \text{if } t_{GTon} + t_{corr} < t < t_{GToff} - t_{corr} \\ \text{NDS}(t) & \text{otherwise} \end{cases} \tag{4.12}$$

The errors are normalized to the total length of the analyzed sentence. Thus, all four single errors (here ERR) are calculated as follows:

$$\text{ERR}_{\%} = \frac{length(\text{ERR})}{length(file)} \tag{4.13}$$

The total error $E_t$of one analyzed sentence i dependent on the selected filter-band, ECM and smoothing method and the selected AD algorithm with or without TCDS, short the combination *comb*, is defined as the sum of the four classified errors:

$$E_t(comb, i) = (\text{FEE}(comb, i) + \text{BEE}(comb, i) + \text{MSE}(comb, i) + \text{NDS}(comb, i) \tag{4.14}$$

The total error of one combination *comb* is defined as the mean of all total errors of all recorded sentences of one user (female or male). In this work, this error is called data set mean total error, as it contains all mean total errors of the evaluation set of the data set of one user.

$$E_t(comb) = \frac{1}{N} \sum_{i=1}^{N} E_t(comb, i) \tag{4.15}$$

The definition of the four timing error areas are presented in figure 4.14. The total time corridor around the on- and offset events $t_{tcorr} = 2t_{corr}$. An example of the detected errors is shown in 4.15.

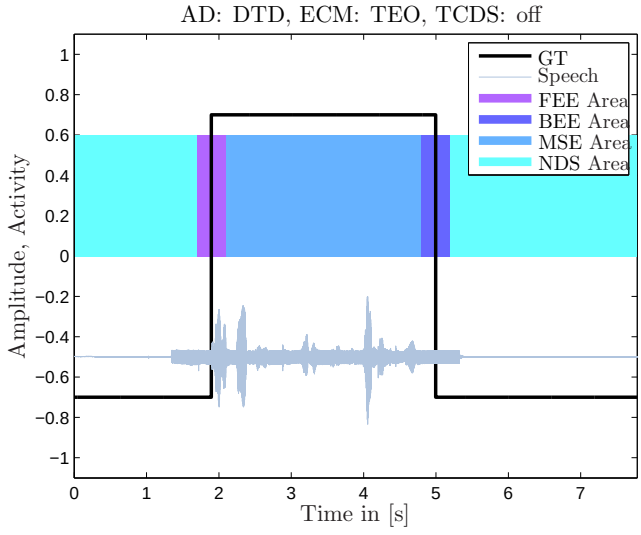


Figure 4.14: Errors regions for FEE, BEE, MSE and NDS of a sentence with a length of 8 s.

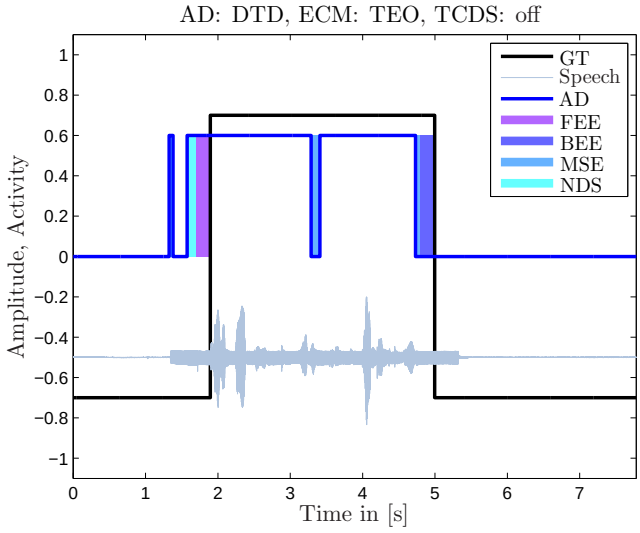


Figure 4.15: Detected Errors after DTD.

4.6.2 Block Detection Ratio

In addition to the time information in the error it is important to find an indicator for the "hackyness", the relation between the correct number of interruptions inside a sentence and the unwanted interruptions due to a wrong behavior of the detection process. The block detection ratio (BDR) is defined as the relation between the number of active blocks in the detection vector NB_{AD} and the desired number of blocks in the GT vector NB_{GT} (equation 4.16). An example of the analyzed blocks of a recorded sentence by a female speaker is given in figure 4.16.

$$\mathrm{BDR}(i) = \frac{\mathrm{NB}_{AD}(i)}{\mathrm{NB}_{GT}(i)} \tag{4.16}$$

The number of detected blocks in the activity detection vector $\mathrm{NB}_{AD} = 4$ and the number of blocks in the GT $\mathrm{NB}_{GT} = 1$. Due to equation 4.16 the BDR of this sentence is 4. A perfect detection in terms of this interruption indicator would result in a BDR of 1. To analyze the total BDR_t of a certain combination *comb* applied on a complete data set the mean of all sentence BDRs is calculated:

$$\mathrm{BDR}_t(comb) = \frac{1}{N}\sum_{i=1}^{N}\mathrm{BDR}(comb, i) \tag{4.17}$$

where N represents the number of analyzed sentences, i represents the sentence number and *comb* represents the used combination of processing methods and AD algorithms.

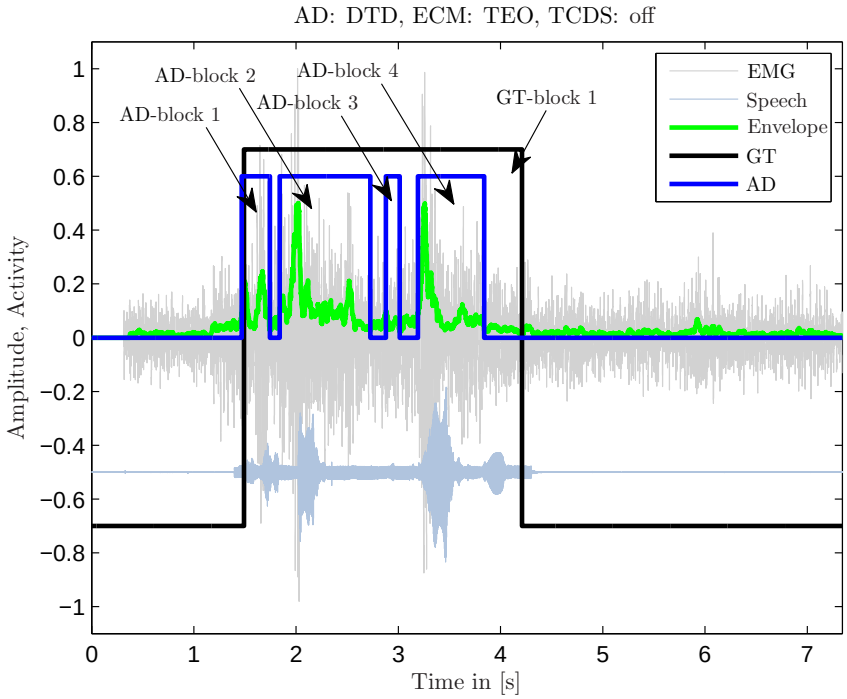


Figure 4.16: Detected Blocks in the AD results compared to the GT block. Smoothing window: 1, Filter-band: 1 (full range), TCDS: off.

5 Results and Discussion

5.1 Digitized sEMG Signal

The recording of the database was done using the amplification and filter stage of the bio-signal shield and a professional audio interface to guarantee best audio quality. Compared to the used sound card (24 bit), the micro-controller system is able to convert the input signal in a resolution of 13 bit, which is enough to perform all processing steps which are proposed in this work without drawbacks in respect to signal detection. The hardware system is designed to work in a real-time scenario and therefor, a smaller amount of data to be processed is an advantage. The sampling rate f_s of the ARDUINO©DUE ADC is set to 8 kHz. This is enough as most of the frequency content of sEMG signals, as explained in section 2.3, is between 0 and 1 kHz.

5.2 On/Off Evaluation

5.2.1 File Structure

The proposed methods are evaluated by applying all combinations mentioned in section 4.3 to the evaluation set of both users. The three AD vectors are calculated and saved to separate files, called *all-on-files*. These files are three dimensional matrices of the format *filename.mat* and the structure $[s \times m \times n]$, where s is the activity vector, consisting of ones and zeros, representing on and off segments, m is the filter band and smoothing selector and n represents the selected detection algorithms and ECMs. Table 5.1 lists the options for m, which can be integers in the range 1 – 28 (4 smoothing windows × 7 filter-bands).

m can be calculated using equation 5.1.

$$m = Settings.band + 8.(Settings.smoothw - 1) - (Settings.smoothw - 1) \tag{5.1}$$

where $Settings.band$ is the selected filter band and $Settings.smoothw$ is the selected smoothing window.

Table 5.1: Filter-band and smoothing selector m.

m	Smoothing Window	Filter-band
1	1 (RMS: 256; HIL, TEO: 512)	1 (0 Hz – 1000 Hz)
2	1 (RMS: 256; HIL, TEO: 512)	2 (0 Hz – 100 Hz)
3	1 (RMS: 256; HIL, TEO: 512)	3 (100 Hz – 200 Hz)
4	1 (RMS: 256; HIL, TEO: 512)	4 (200 Hz – 300 Hz)
5	1 (RMS: 256; HIL, TEO: 512)	5 (300 Hz – 400 Hz)
6	1 (RMS: 256; HIL, TEO: 512)	6 (400 Hz – 500 Hz)
7	1 (RMS: 256; HIL, TEO: 512)	7 (500 Hz – 700 Hz)
8	2 (RMS: 512; HIL, TEO: 1024)	1 (0 Hz – 1000 Hz)
9	2 (RMS: 512; HIL, TEO: 1024)	2 (0 Hz – 100 Hz)
10	2 (RMS: 512; HIL, TEO: 1024)	3 (100 Hz – 200 Hz)
11	2 (RMS: 512; HIL, TEO: 1024)	4 (200 Hz – 300 Hz)
12	2 (RMS: 512; HIL, TEO: 1024)	5 (300 Hz – 400 Hz)
13	2 (RMS: 512; HIL, TEO: 1024)	6 (400 Hz – 500 Hz)
14	2 (RMS: 512; HIL, TEO: 1024)	7 (500 Hz – 700 Hz)
15	3 (RMS: 766; HIL, TEO: 1532)	1 (0 Hz – 1000 Hz)
16	3 (RMS: 766; HIL, TEO: 1532)	2 (0 Hz – 100 Hz)
17	3 (RMS: 766; HIL, TEO: 1532)	3 (100 Hz – 200 Hz)
18	3 (RMS: 766; HIL, TEO: 1532)	4 (200 Hz – 300 Hz)
19	3 (RMS: 766; HIL, TEO: 1532)	5 (300 Hz – 400 Hz)
20	3 (RMS: 766; HIL, TEO: 1532)	6 (400 Hz – 500 Hz)
21	3 (RMS: 766; HIL, TEO: 1532)	7 (500 Hz – 700 Hz)
22	4 (RMS: 1024; HIL, TEO: 2048)	1 (0 Hz – 1000 Hz)
23	4 (RMS: 1024; HIL, TEO: 2048)	2 (0 Hz – 100 Hz)
24	4 (RMS: 1024; HIL, TEO: 2048)	3 (100 Hz – 200 Hz)
25	4 (RMS: 1024; HIL, TEO: 2048)	4 (200 Hz – 300 Hz)
26	4 (RMS: 1024; HIL, TEO: 2048)	5 (300 Hz – 400 Hz)
27	4 (RMS: 1024; HIL, TEO: 2048)	6 (400 Hz – 500 Hz)
28	4 (RMS: 1024; HIL, TEO: 2048)	7 (500 Hz – 700 Hz)

n is a selector, which can hold values from 1 – 24. In the range of 1 – 18, n defines the selected AD algorithm and the used ECM, all with the TCDS option on and off. By choosing n in the range of 19 – 21, the RMS (19), HIL (20) and TEO (21) envelopes of m can be addressed. As the threshold for the ATD is varying over time and not fixed to one value like in the STD and the DTD, this threshold curve depending on m can be addressed by selecting n in the range of 22 – 24, where 22 represents the threshold curve of the analyzed RMS envelope, 23 the HIL and 24 the TEO envelope. Table 5.2 lists all options for n.

In the range of 1 – 18, n can be calculated using equation 5.2.

$$n = 6 * (ad - 1) + 2ecm - 1 + Settings.tcds \tag{5.2}$$

where ad is the selected AD algorithm (STD – 1, DTD – 2, ATD – 3) and ecm is the selected ECM (RMS – 1, HIL – 2, TEO – 3). By choosing $Settings.tcds$ to be 1 or 0, the TCDS algorithm

processed file or the not TCDS processed file can be addressed.

n	AD	ECM
1	STD	RMS
2	STD	RMS + TCDS
3	STD	HIL
4	STD	HIL + TCDS
5	STD	TEO
6	STD	TEO + TCDS
7	DTD	RMS
8	DTD	RMS + TCDS
9	DTD	HIL
10	DTD	HIL + TCDS
11	DTD	TEO
12	DTD	TEO + TCDS

n	AD	ECM
13	ATD	RMS
14	ATD	RMS + TCDS
15	ATD	HIL
16	ATD	HIL + TCDS
17	ATD	TEO
18	ATD	TEO + TCDS
19	RMS envelopes	
20	HIL envelopes	
21	TEO envelope	
22	RMS threshold curves	
23	HIL threshold curves	
24	TEO threshold curves	

Table 5.2: AD and ECM selector n.

Equally to the speech and EMG files, the activity vectors $s(m, n)$, the envelopes as well as the threshold curves for the ATD algorithm are sampled at a sampling frequency f_s of 16 kHz.

5.2.2 Evaluation Results

The evaluation was conducted considering two tuning setups, which are configured to find best results for the AD algorithms under different constraints. In setup 1 (listing 5.1), configured to get best results for STD, the threshold for STD $\mathrm{Th_S}$ and the upper DTD treshold $\mathrm{Th_{Don}}$ are set to 18% of the maximum amplitude value of the envelope. The lower DTD threshold $\mathrm{Th_{Doff}}$ is set to 30% of $\mathrm{Th_{Don}}$. The configuration for ATD in both setups is the same and tuned to achieve best possible results.

```
% STD and DTD parameters:

TH_S = 18;      % in [%]
TH_Don = 18;    % in [%]
TH_Doff = 30;   % in [%] of TH_Don


% ATD Parameters:

Advanced.qmin1 = 0.7;
Advanced.qmin2 = 0.99;
Advanced.qmax1 = 0.7;
Advanced.qmax2 = 0.99;
Advanced.pct = 50;     % Percentage of dynamic range added to current min. power
Advanced.Pdmin = 5;    % Min. dynamics to ignore speech pauses within VAD [dB]
Advanced.tanal = 0.02;  % Analysis times (minimum pause to be analyzed) [s]
```

Listing 5.1: tuning setup 1

In tunings setup 2 (listing 5.2), the parameters are changed to achieve best detection results for DTD. $\mathrm{Th_S}$ and $\mathrm{Th_{Don}}$ are set to 25% of the envelopes maximum value. $\mathrm{Th_{Doff}}$ is set to 50% of $\mathrm{Th_{Don}}$ (12,5% of the envelopes maximum value)

```
% STD and DTD parameters:

TH_S = 25;      % in [%]
TH_Don = 25;    % in [%]
TH_Doff = 50;   % in [%] of TH_Don

% ATD is tuned like in tuning setup 1
```

Listing 5.2: tuning setup 2

The mean of the total error (sum of FEE, BEE, MSE and NDS) of all combinations (the data set mean total error) of user 1 (female speaker) with tuned parameters of tuning setup 1 is shown in figure 5.1. Figure 5.2 shows the same data set mean total error matrix for tuning set 2. The same errors for user 2 (male speaker) are presented in figures 5.3 and 5.4. These figures represent mean total error matrices of all proposed combinations $m \times n$ under two different tuning setups and for both users.

The lightest fields represent low mean total errors of this combination starting from 5%, dark fields refer to higher mean total errors up to 50%. The interpretation and analysis of these matrices is presented in section 5.2.3.

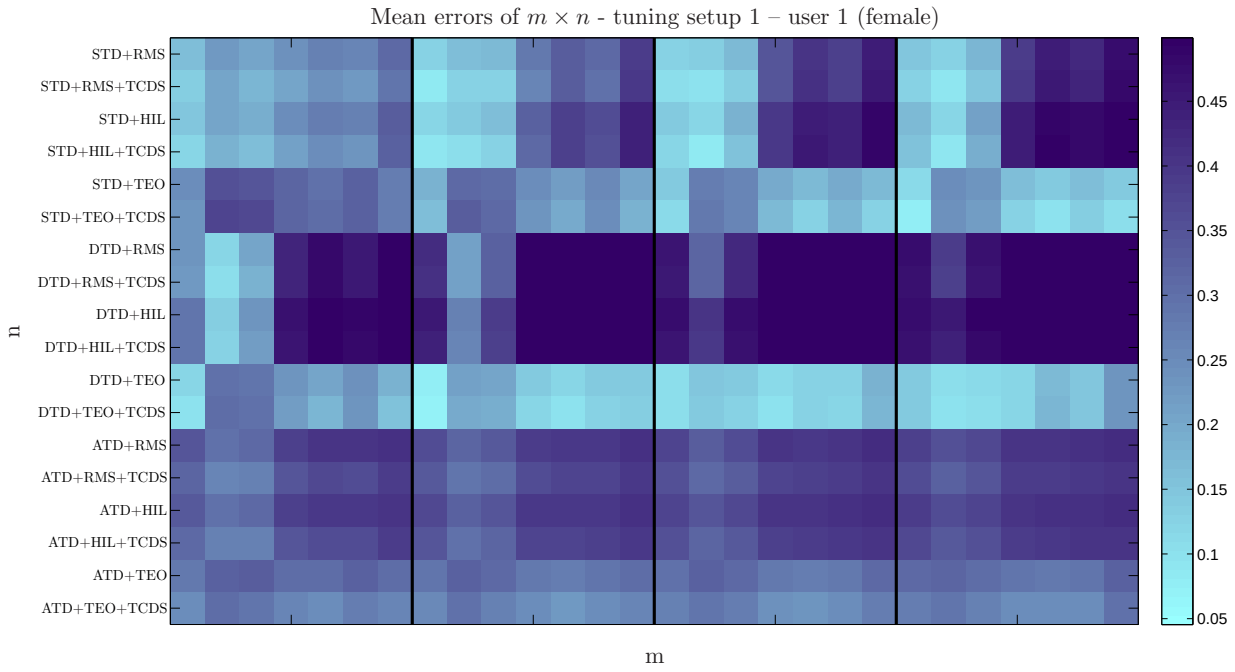


Figure 5.1: Data set mean total error rates of user 1 – Mean of total errors for all combinations – tuning setup 1.

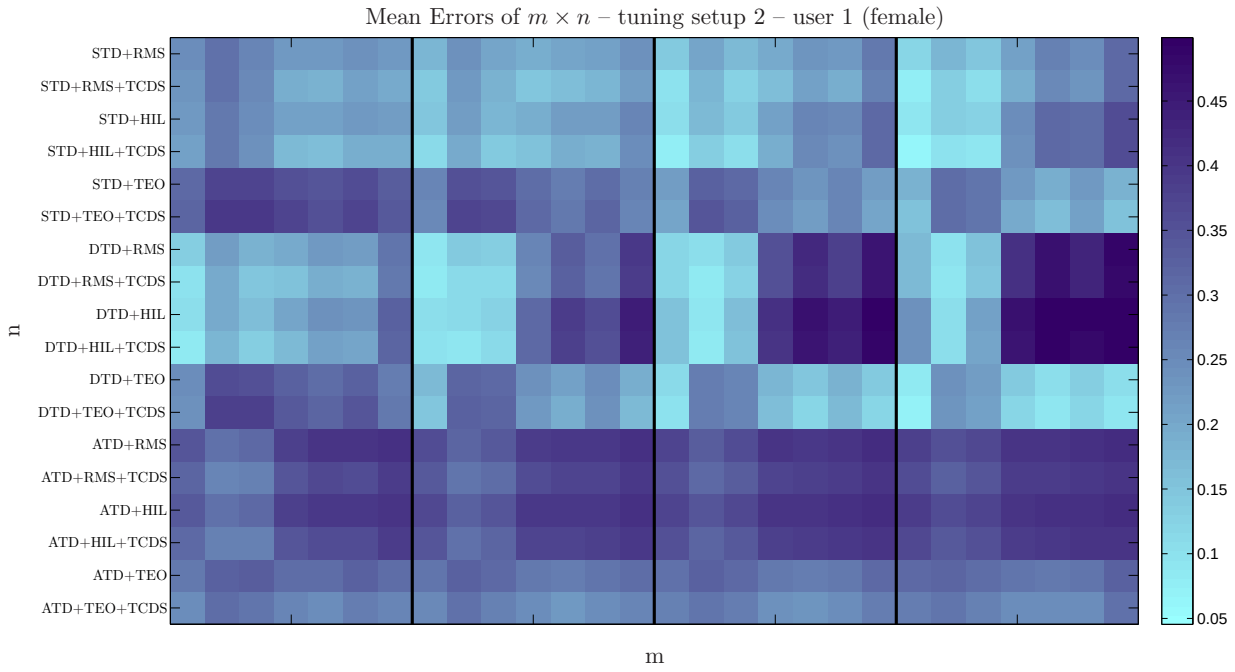


Figure 5.2: Data set mean total error rates of user 1 – tuning setup 1.

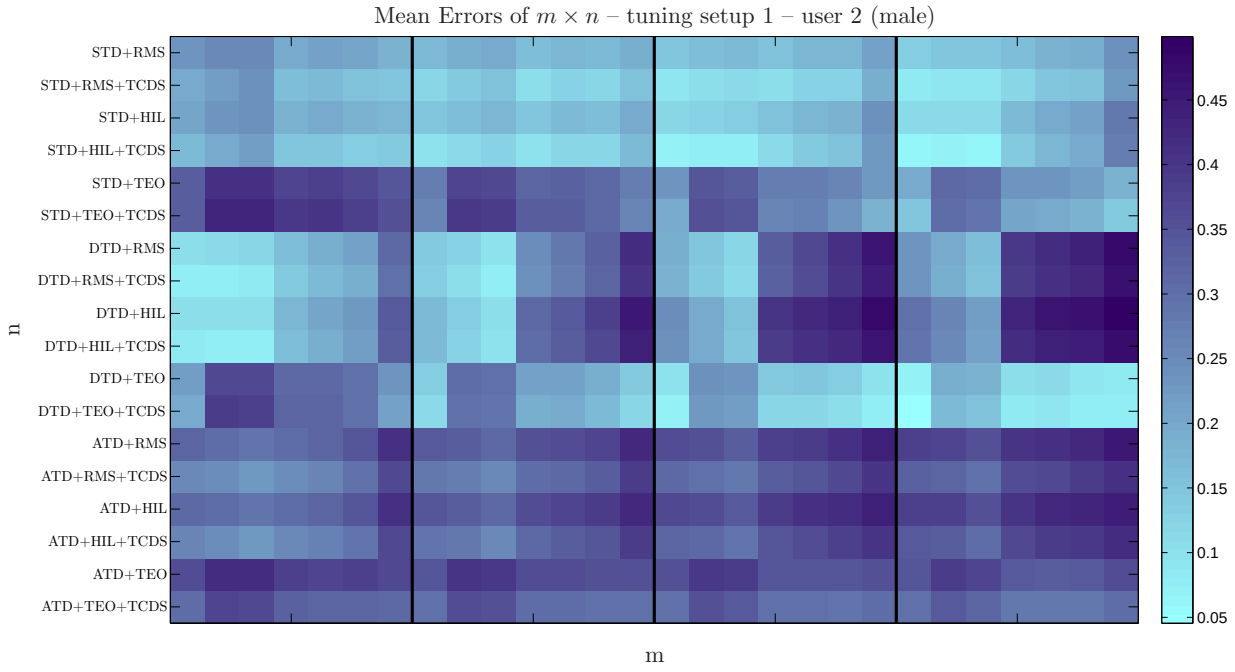


Figure 5.3: Data set mean total error rates of user 2 – tuning setup 1.

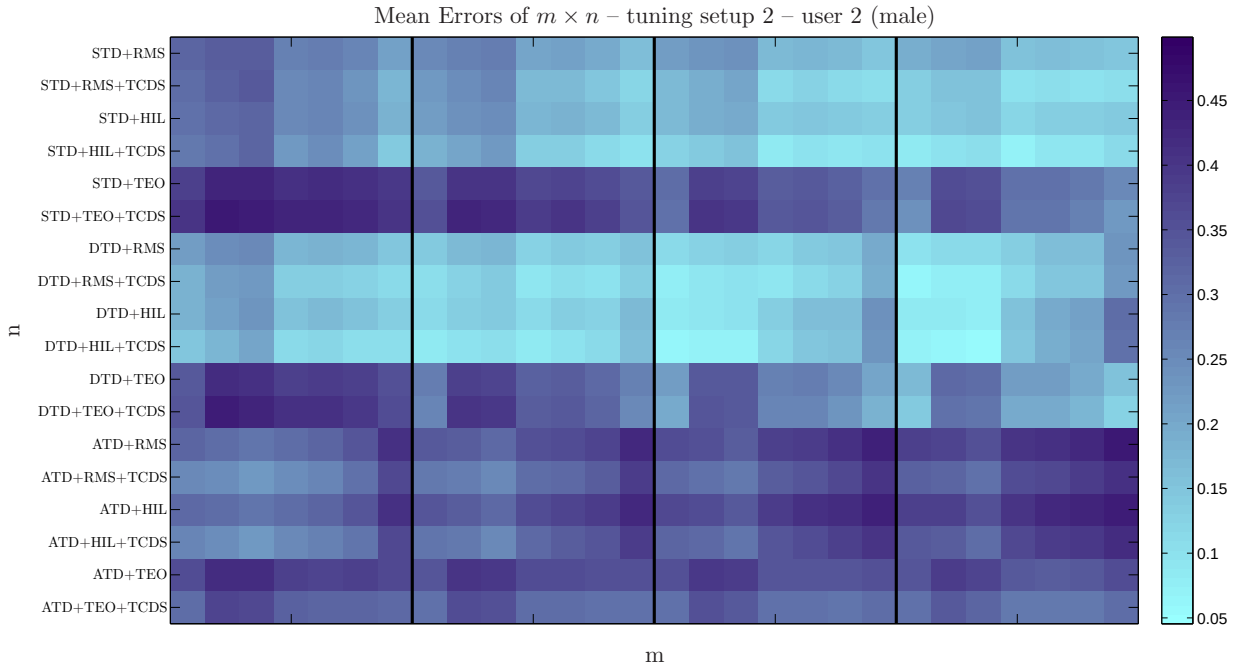


Figure 5.4: Data set mean total error rates of user 2 – tuning setup 2.

5.2.3 User Independent Evaluation

Findings can be separated in user independence and user dependence (section 5.2.4). In the following sections their differences are discussed and significant characteristics of all proposed methods and algorithms are presented.

Time Constant Detection Smoothing

It is evident that the TCDS algorithm, introduced in section 4.5.4, is improving the detection in most cases, except in very few combinations where STD and DTD algorithms are classifying TEO envelopes. Here, the errors are already very high compared to other methods. In two combinations with STD and RMS TCDS is also increasing the error. The maximum decrease of the error of tuning setup 1 and 2 is 5.5 %. The maximum increase is 2.2 % with tuning setup 1 and 2.6 % with tuning setup 2. The mean change of the total error of one method compared to the same method with TCDS with both tuning setups obtains an improvement of the AD. The mean change is a decrease of the error of 2.4 %. Thus, TCDS improves the quality of detection and yields a reduction of the main error in 456 out of 504 combinations which is 90 % of all tested combinations. All results declared in percent are related to the mean total error.

Figures 5.5 and 5.6 show the changes due to TCDS of the mean total errors of all methods and both users of tuning setup 1 and 2. The values are in the range between the maximum increase and the maximum decrease of the error. The change is described as the difference between the mean total error of the combination without TCDS and the combination with TCDS. Blue colored regions indicate an increase of the error between 0 % and 2.2 %, regions colored in red show an improvement and therefor, a decrease of the error higher than 3 %. The change due to TCDS is defined as:

$$\Delta E_t(ad, ecm) = E_t(ad, ecm) - E_t(ad, ecm, tcds) \tag{5.3}$$

where ΔE_t represents the difference between the mean total error of one combination without TCDS and the mean total error of one combination with TCDS.

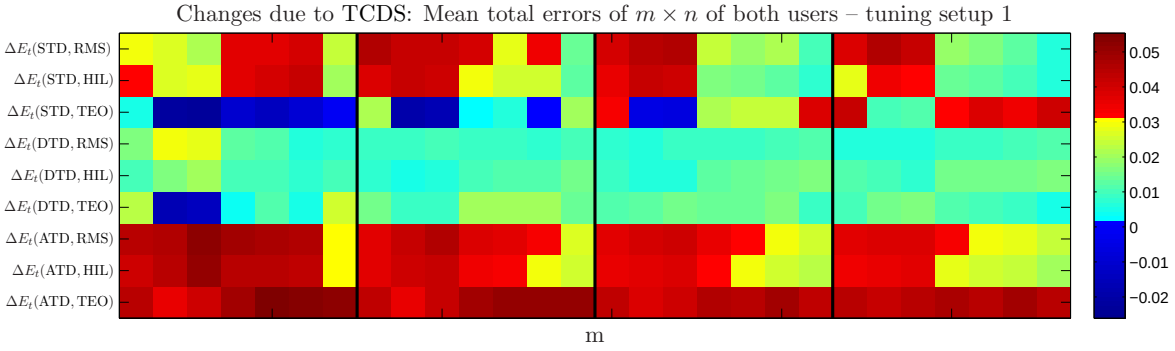


Figure 5.5: Changes in the detection due to TCDS. Mean of total errors for all filter-bands and smoothing windows (m) and all ECMs and AD algorithms (n) tuned with parameters of tuning setup 1.

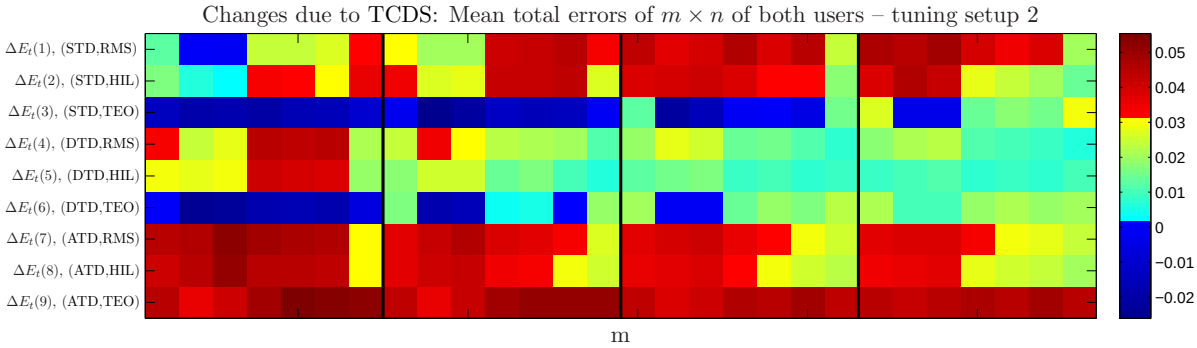


Figure 5.6: Changes in the detection due to TCDS. Mean of total errors for all filter-bands and smoothing windows (m) and all ECMs and AD algorithms (n) tuned with parameters of tuning setup 2.

Threshold Activity Detection Algorithms

In figures 5.1, 5.2, 5.3 and 5.4 it becomes evident that the ATD algorithm does not yield the desired results. Compared to STD and DTD, it shows total error rates between 15 % and 35 %. This algorithm is very powerful when used in an offline setup. Simulated for real-time use, without this smoothing ATD is not providing acceptable results.
In general the total error of RMS and HIL in combination with DTD is increasing with the length of the smoothing window and towards higher filter-bands. Figure 5.7 and 5.8 show this relation for tuning setup 1. It can be seen, that the NDS error is increasing the most. This and the small MSE indicate that the threshold for the DTD is too low and therefor is detecting the signals starting from a too low amplitude. As the thresholds in tuning setup 2 are on a higher level, the MSE becomes lower but still shows an increase with greater smoothing windows and towards higher filter-bands. At the same time, the MSE is decreasing with longer smoothing windows and higher filter-bands (see figures 5.9 and 5.10). It also can be seen that the BDR correlates to the MSE but not to the NDS error. When the MSE is high, the signal is not detected correctly and detection interruptions occur, what leads to an increase of the BDR. In the case where the NDS is high, activity is detected below the SNR of the EMG signal and the probability that all of GT area is detected as well is very high. No increased appearance of interruptions is expected. Thus, no higher BDR will show up.

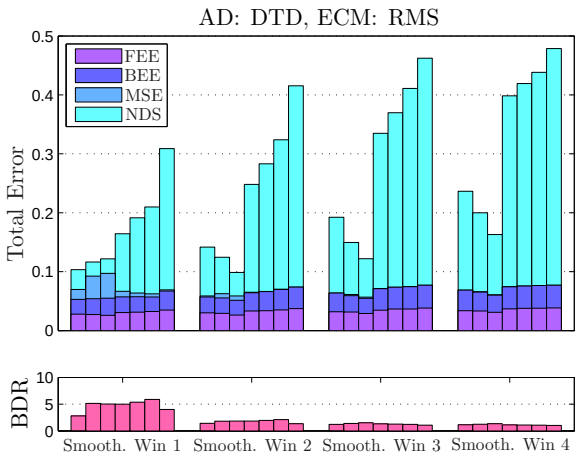


Figure 5.7: Separated errors and BDR of DTD with RMS envelope – tuning setup 1.

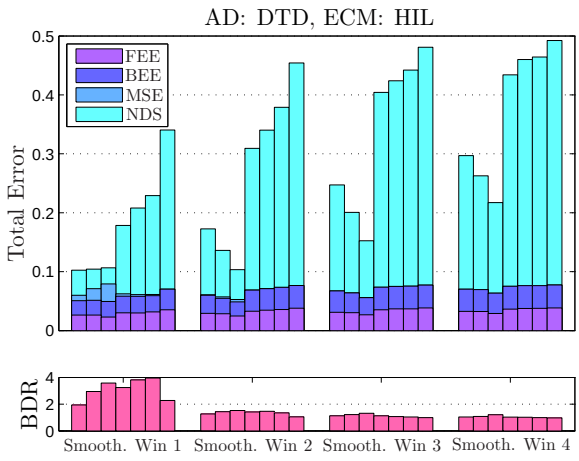


Figure 5.8: Separated errors and BDR of DTD with HIL envelope – tuning setup 1.

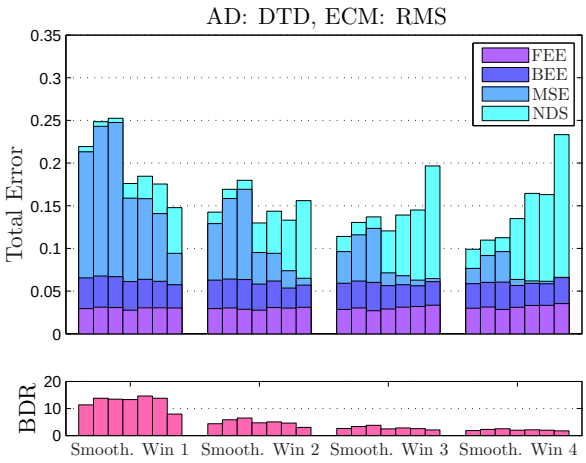


Figure 5.9: Separated errors and BDR of DTD with RMS envelope – tuning setup 2.

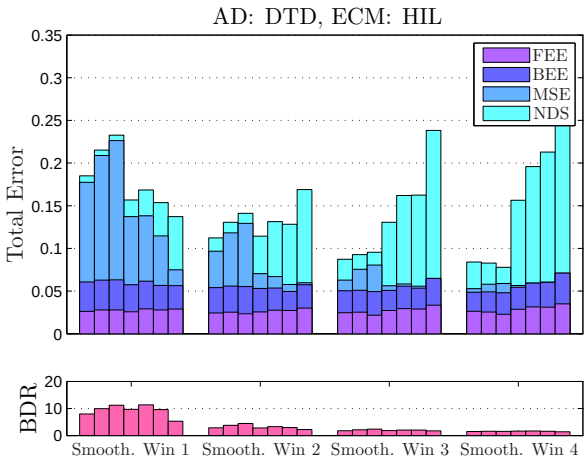


Figure 5.10: Separated errors and BDR of DTD with HIL envelope – tuning setup 2.

Envelope Calculation Methods

The data set mean total error matrices in figures 5.1, 5.2, 5.3 and 5.4 show that there is no significant difference between RMS and HIL in terms of detection. These two ECMs yields nearly the same errors with all AD algorithms. With analyzed RMS and HIL envelopes, all AD algorithms induce an error that is increasing with higher filter-bands and with an increase of the length of the smoothing window except in filter-band 2 (0 Hz – 100 Hz) and 3 (100 Hz – 200 Hz). Using TEO as an input for AD, the error decreases with bigger smoothing window sizes and with higher filter-bands.

5.2.4 User Dependent Evaluation

It has to be said that differences between male and female users in general cannot be specified by the differences of the records of the female and the male user in this work. All findings are interpreted inter-subject but not inter-gender specific.

The most important characteristic in the difference of EMG signals of different users is their SNR. The mean SNR of the recorded user 1 (female) is 11.3 dB and the mean SNR of user 2 (male) is 14.3 dB. In combinations where RMS and HIL amplitudes are used, a higher SNR results in a higher probability of a reliable sEMG detection. Using the TEO as input for the AD, the user's different SNRs yields no significantly differences in the mean errors. Table 5.3 lists the mean errors of all combinations with the three AD algorithms and compares these errors user-specific. The mean errors here are calculated from all smoothing windows. In table 5.4 the mean errors for the three AD algorithms are calculated from all smoothing windows but from filter-bands 1 – 3 only, as it is evident that higher filter-bands yields higher errors in combinations with RMS and HIL.

Table 5.3: User dependent mean errors of combinations with all three ECMs – all filter bands.

AD	Tuning Setup	Mean Error in %		Δ Mean Error in %
		User 1	User 2	
STD	1	25.9	20.7	5.2
DTD	1	35.7	24.2	11.5
ATD	1	33.9	34.4	−0.5
STD	2	23.3	24.4	−1.1
DTD	2	25.3	19.5	5.8
ATD	2	33.9	34.4	−0.5

Table 5.4: User dependent mean errors of combinations with all three ECMs – filter bands 1 – 3.

AD	Tuning Setup	Mean Error in %		Δ Mean Error in %
		User 1	User 2	
STD	1	14.5	14.8	−0.3
DTD	1	35.4	15.5	19.9
ATD	1	33.4	31.2	2.2
STD	2	17.3	22.2	−4.9
DTD	2	13.8	12.8	1.0
ATD	2	33.4	31.2	2.2

General tendencies can be seen here, but as RMS and HIL envelopes cause completely different behaviors of the AD algorithms compared to TEO, it makes sense to exclude TEO from this list.
First of all it can be seen that the mean error gets smaller by selecting the AD-optimized tuning setups for STD and DTD, except for user 1 when all filter-bands are included to calculate the mean error. High error rates in combinations with larger smoothing windows distort the outcome here. The mean errors for the STD are lower in tuning set 1 and the mean errors for the DTD are lower for tuning set 2. The relatively low SNR of user 1 yields a high mean error in combinations with DTD. With higher SNR (user 2), good results can be achieved, although the results depend on the used ECM. Using the parameters of tuning setup 2, the difference between the users can be reduced using the DTD algorithm.

5.2.5 Best Performances

Combinations that offer good performance in terms of detection for both users are selected and described in this chapter. The parameters of all methods and algorithms of the selected combinations as well as the total and the separated mean errors of all sentences of the evaluation set for user 1 and user 2 are presented. Two mean total errors are mentioned in this section. The data set mean total error is defined as the mean of all total errors of all sentences in the evaluation set of the data set of one single user and the user mean total error is defined as the mean of the two data set mean total errors. Therefor, the user mean total error gives information about the performance of the used combination for both users. In most cases much better results can be achieved by finding the best performing combination just for a single user. Furthermore, it

has to be mentioned that the AD algorithms are also detecting non-speech activity, like swallowing, tongue movements or muscle activations due to speech-preparation. These activities are included in $\sim 50\%$ of the sentences in the database. It can be assumed, that better trained users can reduce the NDS error by a significant value and therefor improve the performance of the used combination.

To show the differences and advantages of the selected combinations, box plots (e.g. figure 5.11) are presented. The boxes represent the error values of one combination in which 50% of all sentence errors appear. The height of the box is called the interquartile range (IQR). The dotted lines above and beneath the box, the so-called whiskers, are representing the region where 95% of all sentence total errors appear. The data set mean total error is marked with '×' and the median total error is represented by the middle line of the boxes "notch", which is an indicator of the significance of the change in different boxes. If the region of the notches of two boxes are not overlapping the probability of a significant change is very high. The area of the notch starts at the median value and extends to $\pm 1.58 R_{IQ}/\sqrt{n}$, where R_{IQ} is the IQR and n represents the number of observations [22]. The notch specifies the 95% confidence interval for the difference in two medians. The box colored in pink represents the lowest error rate and therefor marks the combination which offers best detection results.

DTD, TEO + TCDS (combination 1)

In this combination the DTD algorithm is classifying the TEO envelope of the smoothed (smoothing window 2), full-range (filter-band 1) sEMG signal. The on/off information, the result of the tuning setup 1 tuned AD, is smoothed with the TCDS algorithm. All parameters of this combination are presented in table 5.5. Table 5.6 lists total and separated errors. The mean total errors of both users differ by 3.7%, which is the highest percentage compared to the other three combinations presented in this section. This combination gives good results for user 1 but not for user 2 (total mean error > 10%). The user mean total error is 9.15%. Figure 5.11 shows the distribution of the data set mean total error for both users. The pink marked box represents the combination taht results in the lowest possible user mean total error. 50% of all sentence errors are in the range between 3.4% and 11.4%. 95% of all sentence errors are in the range between 0.4% and 23.5%. This wide range indicates a wide distribution and therefor, an unreliable performance for both users.

Table 5.5: Combination parameters

AD	DTD
ECM	TEO
Tuning Setup	1
Smoothing Win.	2
Filter Band	1
TCDS	on

Table 5.6: Separated and total errors of both users.

	User 1	User 2
Data Set Mean Total Error in %	7.3	11.0
User Mean Total Error in %	9.15	
FEE in %	1.4	1.6
BEE in %	2.0	2.0
MSE in %	1.2	6.8
NDS in %	2.7	0.5
BDR	1.5	2.1

This combination results in the worst data set mean total error of the four presented combination in this section, but at the same time provides the best performance for a single user. By choosing smoothing window 4 and filter-band 1, the user mean total error for user 2 is 5%. Thus, this combination is providing the best possible result for a single user of all combinations evaluated in this work. Figure 5.12 shows the distribution of the data set mean total error for user 2. 50% of all sentence errors of the pink colored box are in the range between 2.6% and

6.1%. 95% of all sentence errors are in the region between 0.4% and 10.5%. Compared to all combination, in only one combination (smoothing window 3 and filter band 1) no significant improvement is indicated, considering the overlap of the notches. In all other combinations, a significant decline is indicated. The high smoothing window length (128 ms) could lead to problems because of latency in real-time use.

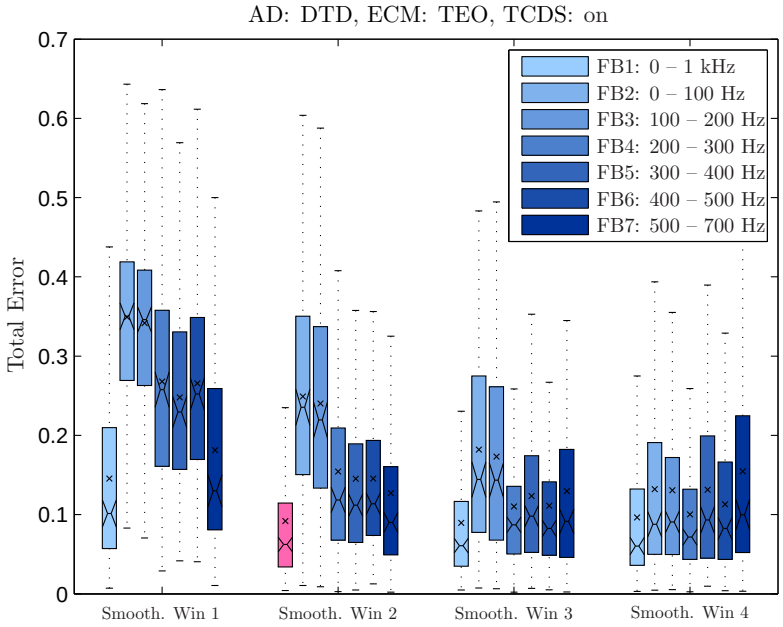


Figure 5.11: Distribution, mean and median total error for evaluation sets of both users.

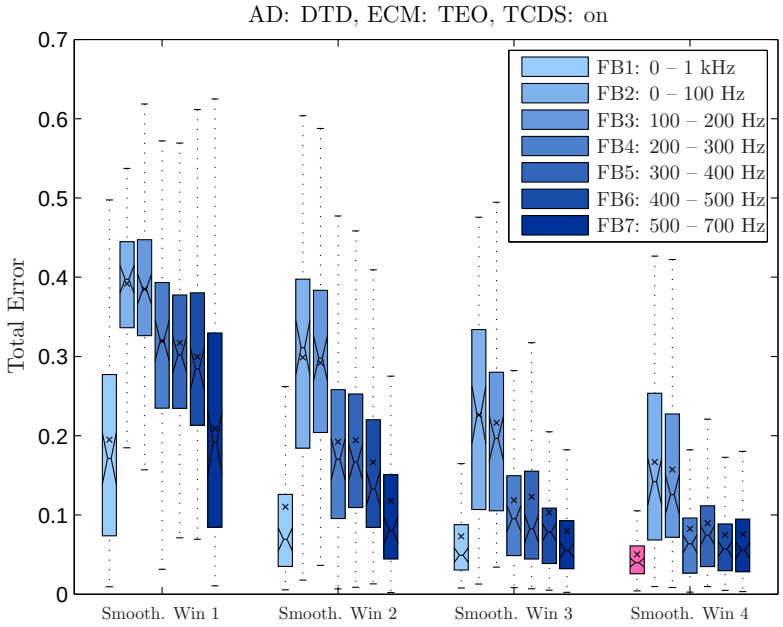


Figure 5.12: Distribution, mean and median total error for evaluation set of user 2.

Figure 5.13 and 5.14 show the separated error distribution for user 1 and user 2. It can be seen that a very low error for in the combination explained above is only showing up in the distribution figure for user 2 (figure 5.14). The same combination for user 1 yields a very high NDS error. The best results for user 1, in this case, is the combination with smoothing window 2 and filter-band 1. The differences in best settings for different users originate in their differences in the SNR. A lower SNR requires smaller smoothing windows to keep the amplitude differences between noise and activity high. To use this combination for on/off triggering in a EL device, a user adjustable smoothing window length could be implemented to select the best working point for the user.

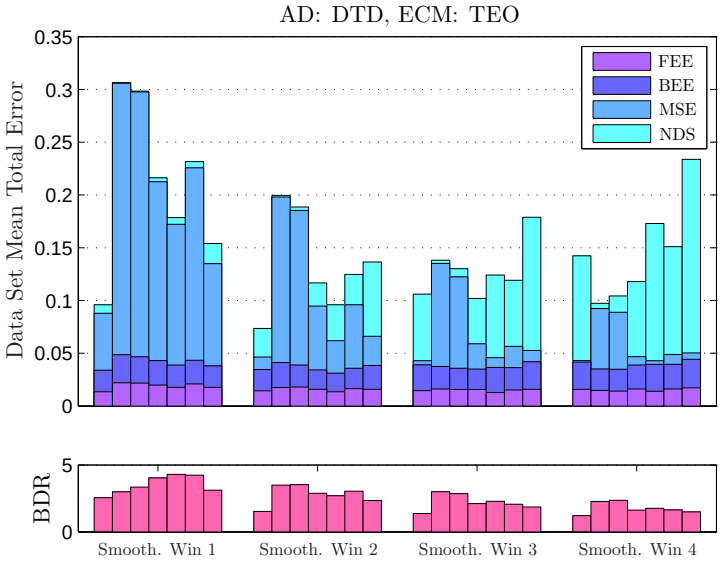


Figure 5.13: Separated errors and BDR of DTD with TEO envelope – tuning setup 1, user 1.

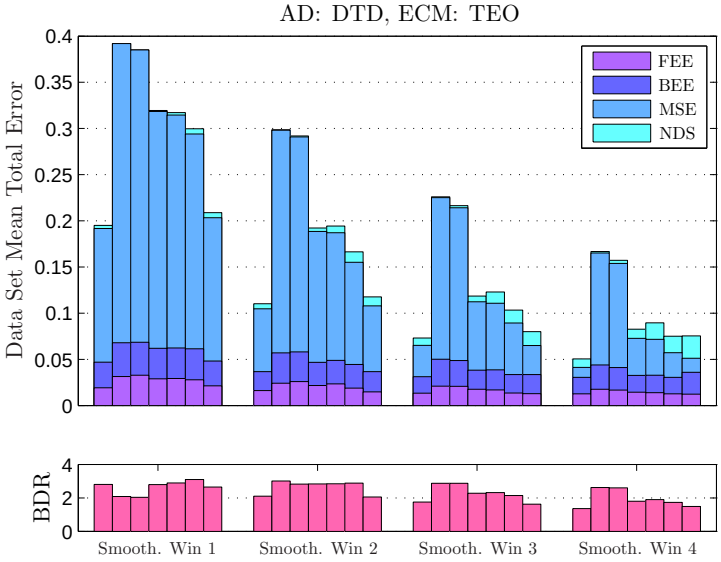


Figure 5.14: Separated errors and BDR of DTD with TEO envelope – tuning setup 1, user 2.

STD, HIL + TCDS (combination 2)

All parameters of this combination are presented in table 5.7. Table 5.8 lists total and separated errors. It shows a similar error tendency of both users, although results for user 1 shows only a bit higher errors. Mainly, the NDS of user 1 is more that twice as high as the NDS of user 2. This indicates that the threshold should be set to a higher value. To guarantee a reliable inter-user functionality, this combination could be realized in an EL device by integrating a user adjustable potentiometer to control the threshold value.

Table 5.7: Combination parameters

AD	STD
ECM	HIL
Tuning Setup	1
Smoothing Win.	3
Filter Band	2
TCDS	on

Table 5.8: Separated and total errors of both users.

	User 1	User 2
Data Set Mean Total Error in %	8.8	7.6
User Mean Total Error in %	8.2	
FEE in %	1.8	1.7
BEE in %	2.0	2.1
MSE in %	2.0	2.5
NDS in %	2.8	1.2
BDR	2.4	1.9

Figure 5.15 shows the distribution of the user mean total error of the selected combinations. In the best combination, 50% of all sentence errors are in the range between 4.5% and 12.1%. 95% of all sentence errors are in the range between 0.5% and 22.3%.

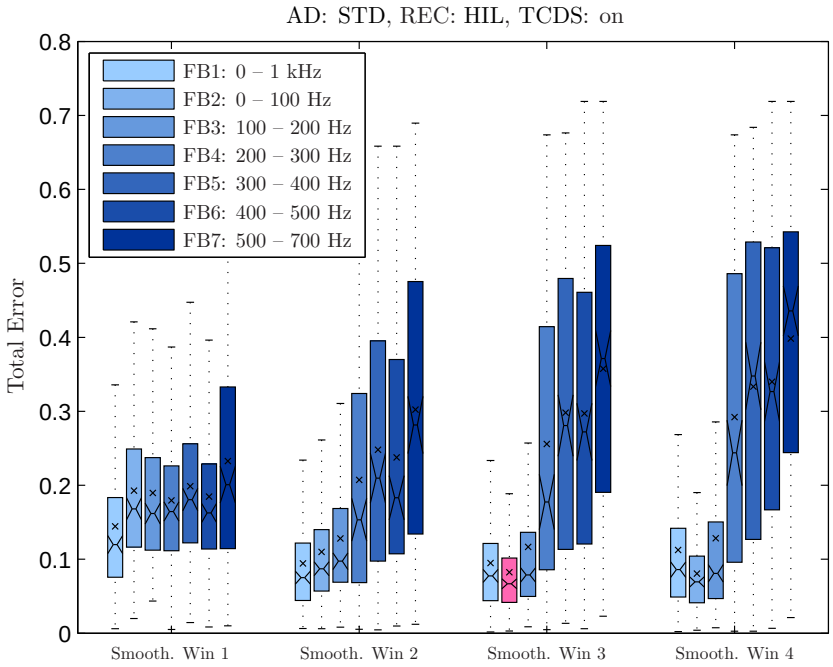


Figure 5.15: Distribution, mean and median total error of all sentences (of evaluation set) and both users.

DTD, RMS + TCDS (combination 3)

All parameters of this combination are presented in table 5.9. Table 5.10 lists total and separated errors. A user mean total error of 8.2% and a BDR of 1.6 indicate that this combination is working very reliably. The results with the selected parameters do not significantly differ from the results with the combination with smoothing window 2 and filter band 1. To reduce latency, this combination should be preferred when selecting a combination for real-time use with an EL device.

Table 5.9: Combination Parameters.

AD	DTD
ECM	RMS
Tuning Setup	2
Smoothing Win.	4
Filter Band	2
TCDS	on

Table 5.10: Separated and total errors of both users.

	User 1	User 2
Data Set Mean Total Error in %	8.8	7.6
User Mean Total Error in %	8.2	
FEE in %	1.8	1.6
BEE in %	2.2	2.1
MSE in %	1.4	2.5
NDS in %	3.4	1.2
BDR	1.6	1.6

Figure 5.16 shows the distribution of the user mean total error. Considering the combination mentioned above, 50% of all sentence errors are in the range between 4.2% and 10.6%. 95% of all sentence errors are in the range between 0.1% and 19.1%.

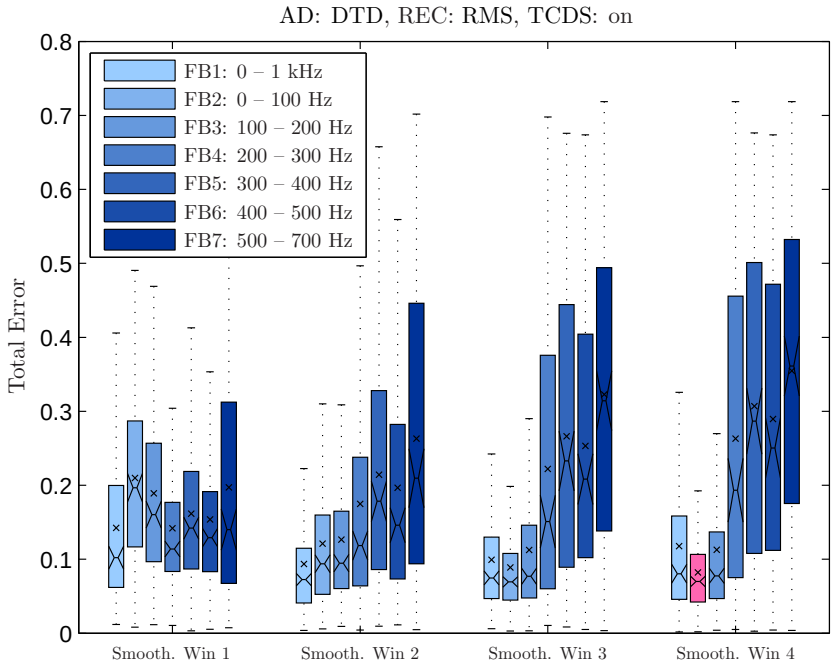


Figure 5.16: Distribution, mean and median total error of all sentences (of evaluation set) and both users.

DTD, HIL + TCDS (combination 4)

Results in this combination only differs marginally from the results in combination 3. All parameters of the combination which is offering the lowest error rate are presented in table 5.11. Table 5.12 lists total and separated errors. The user mean total error of 7.7% shows a 0.5% improvement compared to combination 3. The separated errors show the same tendencies but are slightly lower than in combination 3. Thus it is evident, that DTD in combination with HIL leads to best results for both users. With selecting optimal thresholds for each user even better results may be expected.

Table 5.11: Combination Parameters.

AD	DTD
ECM	HIL
Tuning Setup	2
Smoothing Win.	3
Filter Band	2
TCDS	on

Table 5.12: Separated and total errors of both users.

	User 1	User 2
Data Set Mean Total Error in %	8.7	6.7
User Mean Total Error in %	7.7	
FEE in %	1.7	1.5
BEE in %	2.3	1.8
MSE in %	1.2	2.0
NDS in %	3.5	1.3
BDR	1.6	1.6

Figure 5.17 shows the distribution of the user mean total error. Smoothing window 3 in combination with filter-band 2 yields the lowest errors. 50% of all sentence errors in this combination are in the range between 4.1% and 10.6%. 95% of all sentence errors are in the range between 0.1% and 19.4%.

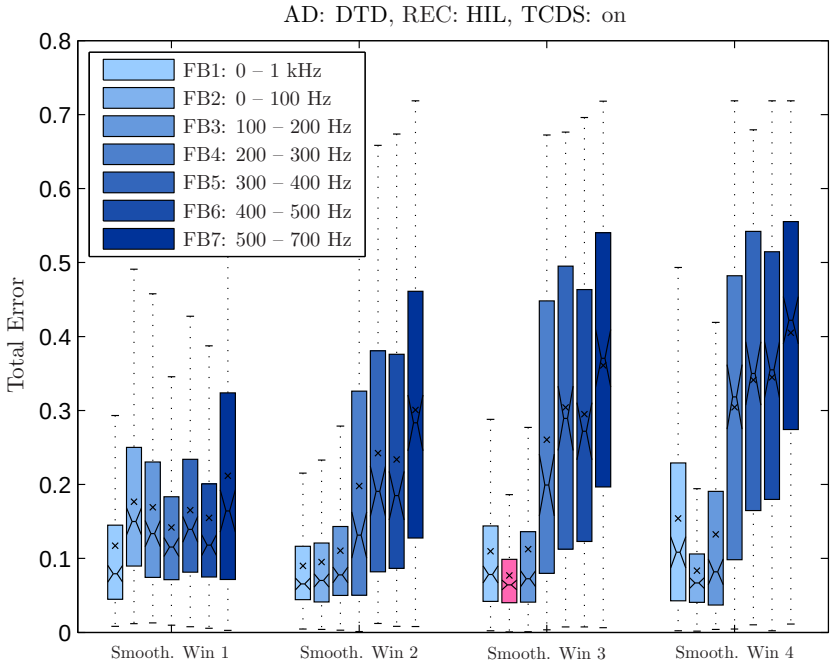


Figure 5.17: Distribution, mean and median total error of all sentences (of evaluation set) and both users.

Threshold determination

As it is not reasonable to calculate every possible parameter configuration in fine steps, and as it was shown that the detection performance depends on the selected threshold, the four combinations presented in section 5.2.5 are analyzed by means of the ROC curves. In ROC curves hit rates $H(thr)$ are plotted against the false alarm rates $F(thr)$ calculated analyzing the database sentences with thresholds thr going from 0% to 100% of the maximal envelope amplitude in steps of 1%. Under the assumptions of signal detection theory interpolating the points for the different values of the threshold leads to a curve. The hit rate and the false alarm rates at a specific threshold are calculated as follows:

$$H(thr) = \frac{\sum_{b=1}^{n_b} \sum_{t=t_{Gon}(b)}^{t_{Goff}(b)} x_{on}(thr,t)}{\sum_{b=1}^{n_b} \sum_{t=t_{Gon}(b)}^{t_{Goff}(b)} GT(thr,t)} \tag{5.4}$$

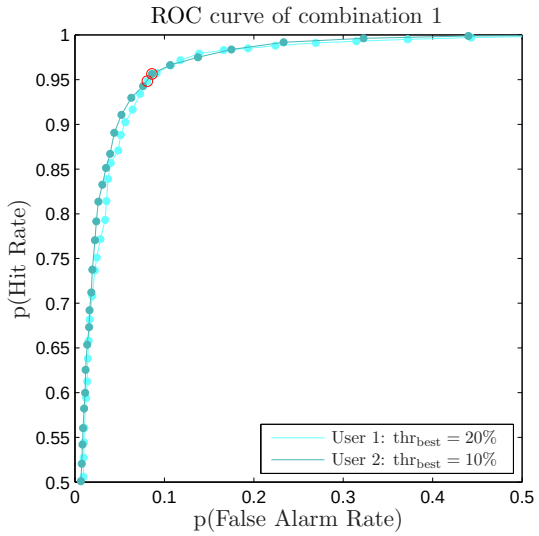
where $x_{on}(thr,t)$ is the detection vector at the actual threshold thr at sample points where the signal is detected, $GT(thr,t)$ is the GT vector of the analyzed sentence at the threshold thr, t_{Gon} and t_{Goff} are the times of the on- and offset of an active GT block, n_b is the number of activity blocks in the GT vector and b is the number of activity blocks of the GT vector.

$$F(thr) = \frac{\sum_{l=0}^{l} x_{off}(thr,l) - \sum_{b=1}^{n_b} \sum_{t=t_{Gon}(b)}^{t_{Goff}(b)} x(thr,t)}{\sum_{l=0}^{l} GT(thr,l) - \sum_{b=1}^{n_b} \sum_{t=t_{Gon}(b)}^{t_{Goff}(b)} GT(thr,t)} \tag{5.5}$$

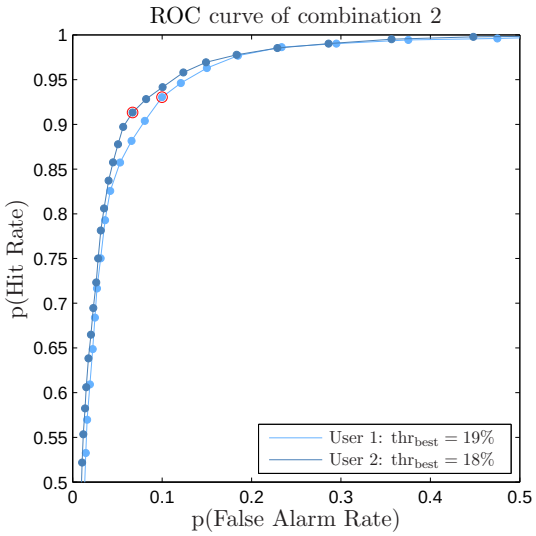
$x(thr,t)$ is the detection vector of the same length like the recorded file, $x_{off}(thr,t)$ is the AD vector at the actual threshold thr at sample points where no activity is detected and l is the length of the recorded sentence.

Sub-figures in figure 5.18 show the ROC curves for the in section 5.2.5 presented combinations. Hit ratios greater that 0.5 and false alarm ratios less than 0.5 are plotted against each other. The single points are linearly interpolated to obtain the curve. The thresholds for best detection (highest hit ratio in combination with lowest false alarm rate) are marked with a red circle and labeled with thr_{best}. These thresholds are calculated by finding the threshold, where the sum of the hit rate and the absolute value of the false alarm rate minus 1 has a maximum.
It can be seen that all algorithms yield better results when the SNR is higher. Thus signals of user 2 are better detected. Furthermore, lower SNRs require higher threshold values for best detection.

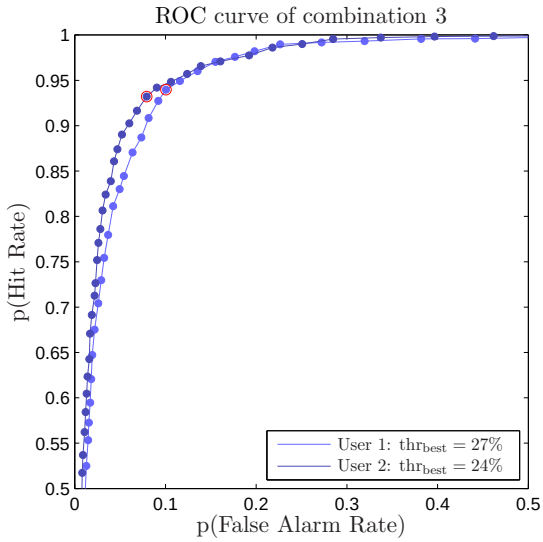
For combination 1, see figure 5.18(a), it becomes clear that the threshold is the determinant factor as the threshold for best detection for user 1 has to be twice as high as the threshold for user 2. As mentioned in section 5.2.5, the smoothing window also effects the quality of detection between users very much. Thus in the real-time implementation, a user adjustable threshold and smoothing window have to be provided. In combination 2, see figure 5.18(b), the thresholds for best detection for both users are very close. This fact indicates that in combination 2, it is not necessary to adjust the threshold for different users. In combination 3, presented in figure 5.18(c), as well as in combination 2 the false alarm rate for user 2 at the best detection threshold is very close to 10%, which is not satisfying. In respect of the ROC analysis combination 1 and 4 yields the best results. Thus ROC analysis is confirming the findings in section 5.2.5. In these combinations either one or two adjustable parameters (threshold and smoothing window) should be implemented in the real-time EL device.



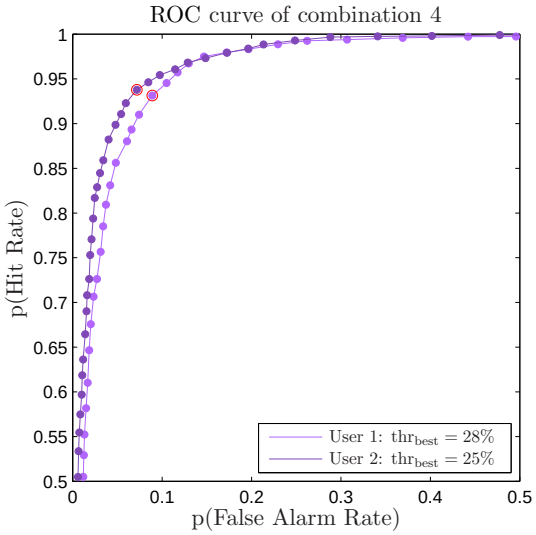
(a) ROC curve of combination 1, $H(thr)$ vs. $F(thr)$ of user 1 and user 2.



(b) ROC curve of combination 2, $H(thr)$ vs. $F(thr)$ of user 1 and user 2.



(c) ROC curve of combination 3, $H(thr)$ vs. $F(thr)$ of user 1 and user 2.



(d) ROC curve of combination 4, $H(thr)$ vs. $F(thr)$ of user 1 and user 2.

Figure 5.18: ROC curves of the four combinations presented in section 5.2.5 for both users.

5.2.6 Error Interpretation

User dependent and independent error rates are introduced to get result indicators which are close to subjective user perception. By definition, FEE and BEE rates are expected to be lower than MSE and NDS rates. These error rates define detection errors at transition times in the GT. Nevertheless, FEE and BEE rates higher than 2% indicate that the activity on- or offset is not detected properly. MSE and NDS rates higher than 2% indicate more serious detection errors, as this errors occur where no transitions in the GT are defined. The detection of interruptions during GT activity yields an increase in the BDR and in many cases goes along with an increase of the MSE rate. Whether these interruptions lead to a discomfort in the real-time usage has to be evaluated in a user study. If the interruptions occur during short unvoiced periods, it can be assumed that these interruptions are not affecting the articulation or even are improving it.

The results of the ROC analysis is not as close at the users perception than the user dependent and independent errors. In general, the shape of the curve indicate if the combination offers reliable detection. The closer the curve approaches the upper left corner, the better the detection is. In the upper left corner the hit rate is 1 and the false alarm rate is 0. Here, 100% of the signal is detected, while 0% of the noise is detected as signal. The results of the ROC analysis also show how well a combination can be used for users with different SNRs. Small differences between the optimal thresholds for different users indicate that the combination shows low impacts due to different SNRs.

6 Conclusions

6.1 Signal Acquisition

This work consists of two main parts: The development of the signal acquisition hardware and the proposed signal processing and activity detection (AD) methods. The bio-signal shield is a portable, high quality, easy-to-use surface electromyography (sEMG) amplification and filter circuit, compatible with the popular ARDUINO©DUE micro-controller board, which is used for analog-digital conversion of the signal. This compatibility, the portable design and the wide, adjustable amplification range make this system competitive to expensive diagnostic systems in many application areas. The stackable design of the shield offers the possibility to connect up to six sensor units and therefor to record up to six sEMG channels at the same time. This device provides an integrated half-wave splitter to increase the amplitude resolution of the digitized signal to 13 bit. It can serve as a bio-signal recording tool and as a development system for live applications in which sEMG signals are used for control purposes.

6.2 Proposed Method

The recorded database, a phonetically balanced set of 114 sentences, spoken by two healthy speakers provides the basis for the evaluation of different electromyography (EMG) signal processing and classification methods. The signal is cleaned from interferences caused by the excitation signal transmitted by an electrolarynx (EL) prototype using adaptive noise cancellation (ANC). The enhanced signal is split into sub-bands and root mean square (RMS), Hilbert transform (HIL) and Teager energy operator (TEO) envelopes of these signals are calculated. The envelopes are smoothed using a moving average (MA) filter with different window lengths. These envelopes are used as inputs to three activity detection (AD) algorithms. The single threshold detection (STD), the double threshold detection (DTD) and the adaptive threshold detection (ATD) algorithm are classifying the signal in terms of activity. The detection results are compared to speech activity ground truth (GT), which is manually defined by analyzing the corresponding speech data. Time-dependent and time-independent error rates are calculated and reliable combinations of sub-band ranges, smoothing window lengths, envelope calculation methods (ECMs) and classification methods are evaluated. A detection smoothing algorithm for the use in live applications is presented which increases the detection rate by 2.4% on average compared to the not smoothed results of the AD algorithms. Using receiver operating charac-

teristics (ROC) analysis, optimal working points are determined. Best configurations result in user mean total error rates between 7.7% and 9.15%.

6.3 Outlook

The integration of the proposed system into a real-time EL system is the next step. Results of a subjective user test in which the proposed configurations are evaluated should be compared to the findings in this work. Analyzing the accordance of the findings in this work and the results of the user study would gather important insights about how well the introduced error rate calculation fits the subjective perception. To improve the detection rate, a system to identify non-speech related muscle activity, like swallowing may be implemented. In a next step, varying pitch information can be extracted from the sEMG signal to overcome the second large drawback of conventional EL systems, its monotonic sound. An evaluation system including a user test has to be developed and its results have to be evaluated. This could result in an EL prototype system with an automatic on/off switch and a natural sounding pitch contour. Finally, a commercial, fully integrated, portable, neck-mounted EL system could be developed.

Acronyms

AD Activity detection . . . 17

ADC Analog-digital converter . . . 12

ANC Adaptive noise cancellation . . . 17

ALE Adaptive line enhancement . . . 20

ATD Adaptive threshold detection . . . 17

BDR Block detection ratio . . . 29

BEE Back end error . . . 27

CMRR Common-mode rejection ratio . . . 7

CPU Central processing unit . . . 8

DC Direct current . . . 10

DIP Dual in-line package . . . 9

DRL Driven-right-leg . . . 7

DTD Double threshold detection . . . 6

ECG Electrocardiography . . . 10

ECM Envelope calculation method . . . 17

EL Electrolarynx . . . 1

EMG Electromyography . . . 1

FEE Front end error . . . 27

FIR Finite impulse response . . . 22

GT Ground truth . . . 17

GMM Gaussian mixture Model . . . 6

HIL Hilbert transform 17

IC Integrated circuit 9

IQR Interquartile range 41

LF Liljencrants–Fant 18

MA Moving average 18

MSE Middle speech error 27

MUAP Motor unit action potential 4

NDS Noise detected as speech 27

PCB Printed circuit board 2

RF Radio frequency 6

RMS Root mean square 17

ROC Receiver operating characteristics 2

sEMG Surface electromyography 2

SMD Surface-mount device 9

SNR Signal-to-noise ratio 6

SOIC Small outline integrated circuit 9

STD Single threshold detection 6

SVM Support vector machine 6

TCDS Time constant detection smoothing 27

TEO Teager energy operator 17

TP Test point 11

USB Universal serial bus 8

VAD Voice activity detection 26

A Appendix

A.1 Hardware Development

A.1.1 Bio-signal Shield Circuit Design

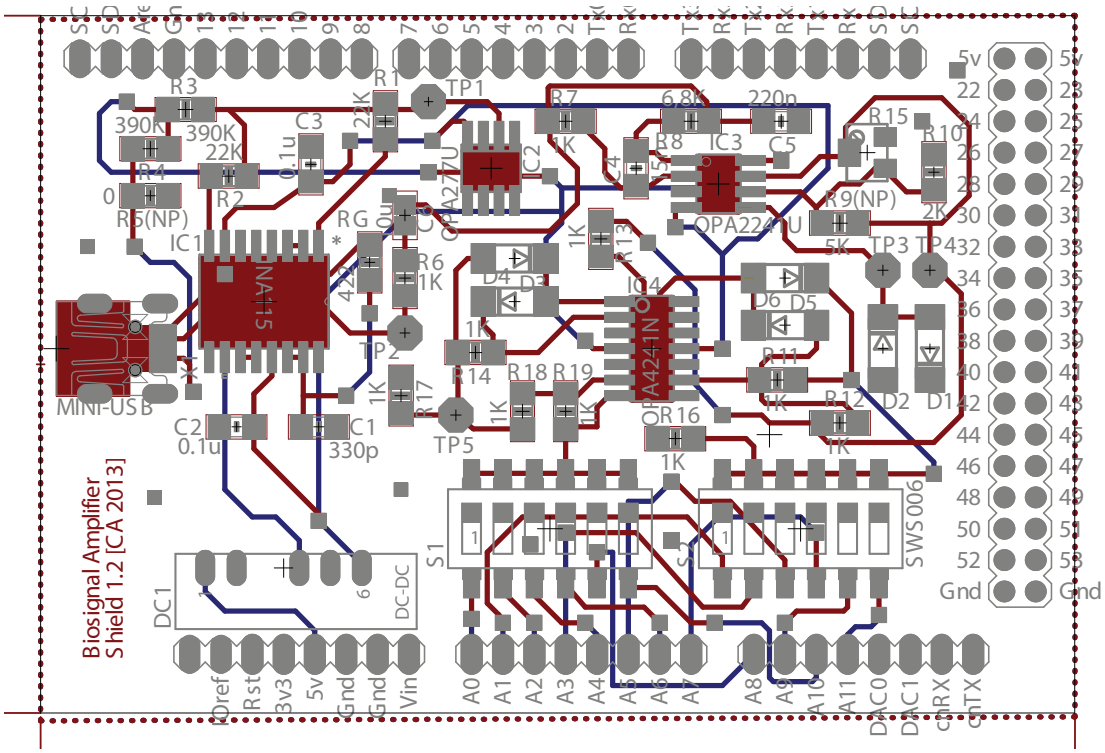


Figure A.1: PCB layout of the bio-signal shield.

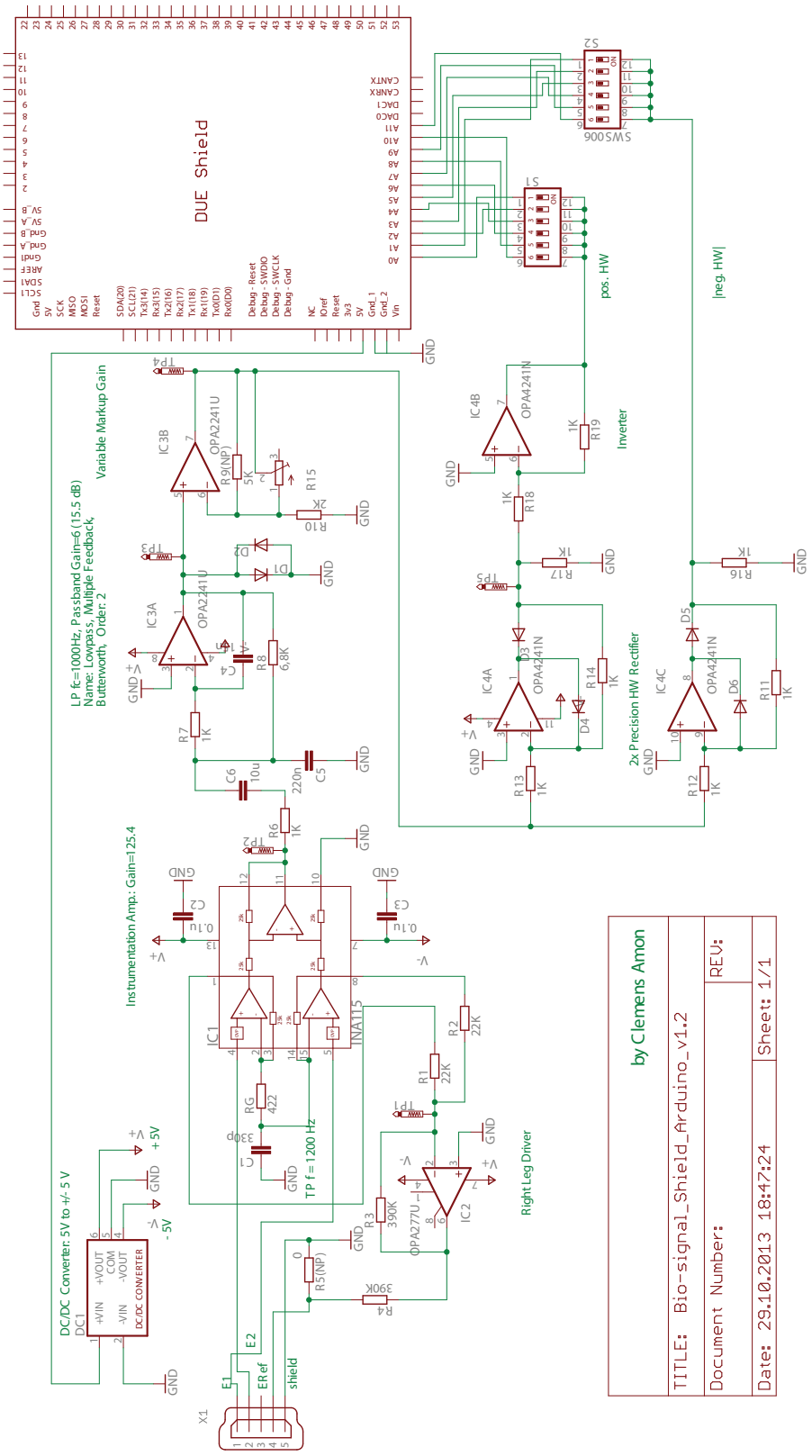


Figure A.2: Electrical circuit of the bio-signal shield version 1.2.

A.1.2 Version Changelog

Changes from bio-signal shield version 1.1 to 1.2

- Adding the missing connection between resistors R_{12} and R_{13}
- Adding a via near resistor R_{15} to expand ground to this area
- Adding a via below the switch S1 to expand ground to this area
- The capacitor C_6s was moved to the input of the operational amplification stage to avoid the amplification of an undesired DC offset, which is eliminated by this capacitor.

A.2 EMG and Speech Database

The recorded database consists of 114 phonetically balanced sentences in German language split into two data sets: The development set, presented in table A.1 and the evaluation set, presented in table A.2.

Table A.1: Development Set

No.	Sentence
NS00102	Sie wurden einig, dass derjenige für den Stärkeren gelten sollte, der den Wanderer zwingen würde, seinen Mantel abzunehmen.
sen0004	Gestern stürmte es noch.
sen0015	Hans isst so gerne Wurst.
sen0027	Jetzt suche ich das Weißbrot.
sen0045	Die Kartoffeln gehören zum Mittagessen.
sen0060	Am Zaun steht eine Regentonne.
sen0067	Wohin fährt dieser Zug?
sen0081	Die KATZE jagt eine Maus
sen0094	Schon bald sind wir zu Hause.

Table A.2: Evaluation Set

No.	Sentence
NS00101	Einst stritten sich Nordwind und Sonne, wer von ihnen beiden wohl der Stärkere wäre, als ein Wanderer, der in einen warmen Mantel gehüllt war, des Weges daherkam.
NS00103	Der Nordwind blies mit aller Macht, aber je mehr er blies, desto fester hüllte sich der Wanderer in seinen Mantel ein.
NS00104	Endlich gab der Nordwind den Kampf auf.
NS00105	Nun erwärmte die Sonne die Luft mit ihren freundlichen Strahlen, und schon nach wenigen Augenblicken zog der Wanderer seinen Mantel aus.
NS00106	Da musste der Nordwind zugeben, dass die Sonne von ihnen beiden der Stärkere war.
sen0000	Heute ist schönes Frühlingswetter.
sen0001	Die Sonne lacht.
sen0002	Am blauen Himmel ziehen die Wolken.

sen0003 Über die Felder weht ein Wind.
sen0005 Montag war es uns zu regnerisch.
sen0006 Ich werde mit der Fähre nach Irland übersetzen.
sen0007 Ich werde den Text ins Englische übersetzen.
sen0008 Der BÄR hat den Fisch gefangen.
sen0009 Der Bär hat den FISCH gefangen.
sen0010 Der Bär hat den Fisch GEFANGEN.
sen0011 Mutter konnte länger schlafen.
sen0012 Der Kaffee dampft in den Tassen.
sen0013 Messer und Gabel liegen neben dem Teller.
sen0014 In der Mitte steht der Brötchenkorb.
sen0016 Gib mir bitte die Butter!
sen0017 Wer möchte noch Milch?
sen0018 Bald ist der Hunger gestillt.
sen0019 Günther muss noch einkaufen gehen.
sen0020 Achte auf die Autos!
sen0021 Überquere die Straße vorsichtig!
sen0022 Sonst wirst du leicht überfahren.
sen0023 Radfahrer sausen vorbei.
sen0024 Im Geschäft stehen viele Leute.
sen0025 Gleich hier sind die Nahrungsmittel.
sen0026 Muss der Zucker nicht dort drüben stehen?
sen0028 Ob ich Süßigkeiten kaufen darf?
sen0029 Hier gibt es Konserven.
sen0030 Die Oma trinkt einen Kaffee.
sen0031 Die Oma trinkt einen Kaffee?
sen0032 Seine Frau macht ein trauriges Gesicht.
sen0033 Du solltest weniger rauchen.
sen0034 Die Ärzte sind damit gar nicht einverstanden.
sen0035 Gib mir bitte mal die Zeitung!
sen0036 Aber Schönes steht wohl nicht drin.
sen0037 Wer muss noch Schularbeiten machen?
sen0038 Ich müsste lesen und rechnen.
sen0039 Sieglinde zeichnet eine Figur.
sen0040 Der Opa fährt ein blaues Fahrrad.
sen0041 Der Opa fährt ein blaues Fahrrad?
sen0042 Wir wollen heute spazieren gehen.
sen0043 Da möchte ich gerne mit.
sen0044 Zuvor müssen wir uns stärken.
sen0046 Zum Schnitzel gibt es Erbsen.
sen0047 Dazu essen wir den Salat.
sen0048 Wer trinkt einen Kaffee?
sen0049 Danach tut eine Wanderung gut.
sen0050 Können wir nicht Tante Erna besuchen?
sen0051 Zieht vielleicht die festen Schuhe an!
sen0052 Zurück geht's mit der Bahn.
sen0053 Durch Wald und Feld führt unser Weg.
sen0054 Wir hören den plätschernden Bach.
sen0055 Hasen verschwinden im Dickicht.
sen0056 Voller Glück sind wir am Ziel.

sen0057	Die Tante bewohnt ein nettes Häuschen.
sen0058	Dahinter liegt der Rosengarten.
sen0059	Manche Obstbaeume blühen prächtig.
sen0061	Der gelbe Küchenofen sorgt für Wärme.
sen0062	Im Topf kocht das Wasser.
sen0063	Ein Sofa steht an der Wand.
sen0064	In welche Richtung geht es zum Hauptplatz?
sen0065	Wer ist für dieses Chaos verantwortlich?
sen0066	Kann mir niemand sagen, was hier eigentlich los ist?
sen0068	Wann kommt endlich das Taxi zum Flughafen?
sen0069	Wann wird es endlich wärmer?
sen0070	Geh mir bitte aus dem Weg!
sen0071	Gib mir die Bohrmaschine!
sen0072	Die drei Männer sind begeistert.
sen0073	Vater mischt gleich die Karten.
sen0074	Er gewinnt sechs Spiele nacheinander.
sen0075	Ist es nicht Zeit zum Aufbruch?
sen0076	Der Bahnhof liegt sieben Minuten entfernt.
sen0077	Löst doch die Fahrkarten am Schalter!
sen0078	Wir gehen auf den Bahnsteig.
sen0079	Da läuft der Zug ein.
sen0080	Die Bremsen quietschen gräßlich.
sen0082	Die Katze JAGT eine Maus
sen0083	Die Katze jagt eine MAUS
sen0084	Wir haben ein Abteil extra für uns.
sen0085	Der junge Zugbegleiter pfeift zur Abfahrt.
sen0086	Leise rollen wir aus dem Bahnhof.
sen0087	Draußen fliegt die Landschaft vorbei.
sen0088	Die Rinder sind noch auf der Weide.
sen0089	Ein Bauer arbeitet auf seinem Acker.
sen0090	Der Pflug zieht tiefe Furchen.
sen0091	Daneben grünt schon Wintersaat.
sen0092	Hier richten Zimmerleute ein Dach.
sen0093	Es gehört zu einer Feldscheune.
sen0095	Die Fahrt war ja mächtig kurz.
sen0096	Zug endet hier! verkündet die Ansage.
sen0097	Alle eilen gleich links ins Freie.
sen0098	In der Dämmerung kommen wir heim.
sen0099	Das war jetzt aber ein schöner Tag.

Bibliography

[1] F.W. Akin, O.D. Murnane, J.W. Tampas, and C.G. Clinard. The effect of age on the vestibular evoked myogenic potential and sternocleidomastoid muscle tonic electromyogram level. *Journal of the American Auditory Society - Ear and Hearing*, Sep. 2011.

[2] J.L. Andreassi. *Psychophysiology: Human Behavior and Physiological Response*. Lawrence Erlbaum Associates, 2000.

[3] J.E. Atkinson. Correlation analysis of the physiological factors controlling fundamental voice frequency. *Journal of the Acoustican Society of America*, pages pp. 211–222, Jan. 1978.

[4] J.V. Basmajian and C.J. de Luca. *Muscles Alive - The Functions Revealed by Electromyography*. Williams and Wilkins Company, Baltimore, 1985.

[5] F. Beritelli, S. Casale, and G. Ruggeri. Performance evaluation and comparison of itu-t/etsi voice activity detectors. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICAASP)*, volume 3, pages 1425–1428, 2001.

[6] Burr-Brown. Precision Instrumentation Amplifier INA115 - Datasheet. Technical report, Texas Instruments, 2013.

[7] OKAWA Electric Design. Engineering design utilities - online filter design resource. *http://sim.okawa-denshi.jp/en/*, 2008.

[8] Scott C. Douglas. Adaptive line enhancer (ALE) - Matlab DSP system toolbox example. *http://www.mathworks.de/de/help/dsp/examples/adaptive-line-enhancer-ale.html*, 2013.

[9] C. Draxler and K. Jänisch. Speechrecorder - A universal platform independent multichannel audio recording software. *Proceedings of the Fourth International Conference on Language Resources and Evaluation Lisbon, Portugal, pp. 559-562*, 2004.

[10] G. Fant, J. Liljencrants, and Q. Lin. A four-parameter model of glottal flow. *Speech Transmission Laboratory - Quarterly Progress and Status Report*, pages 1–13, 1985.

[11] E.A. Goldstein, J.T. Heaton, J.B. Kobler, G.B. Stanley, and R.E. Hillman. Design and implementation of a hands-free electrolarynx device controlled by neck strap muscle electromyographic activity. *IEEE Transactions on Biomedical Engineering*, 51:325–332, 2004.

[12] E. Hansler and Gerhard Schmidt. *Acoustic Echo and Noise Control: A Practical Approach*. Wiley-IEEE Press, 1 Edition, 2004.

[13] James T. Heaton, Mark Robertson, and Cliff Griffin. Development of a wireless electromyographically controlled electrolarynx voice prosthesis. *Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society.*, 2011:5352–5, Jan. 2011.

[14] C. Iber, S. Ancoli-Israel, A. Chesson, and S.F. Quan. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specification*. Westchester, Illinois: American Academy of Speep Medicine, 2007.

[15] Markus Iseli and Abeer Alwan. Inter- and intra-speaker variability of glottal flow derivative using the lf model. In *Proceedings of Annual Conference of the International Speech Communication Association (ISCA)*, pages 477–480, 2000.

[16] Matthias Janke. Spektrale Methoden zur EMG-basierten Erkennung lautloser Sprache. Master's thesis, Cognitive Systems Lab, Karlsruhe Institute of Technology (KIT), 2010.

[17] Matthias Janke, Michael Wand, Keigo Nakamura, and Tanja Schultz. Further investigations on EMG-to-speech conversion. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 365–368. Cognitive Systems Lab, Karlsruhe Institute of Technology (KIT), 2012.

[18] J.F. Kaiser. Some useful properties of Teager's energy operators. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICAASP)*, volume 3, pages 149–152, 1993.

[19] Heather L. Kubert, Cara E. Stepp, Steven M. Zeitels, John E. Gooey, Michael J. Walsh, S.R. Prakash, Robert E. Hillman, and James T. Heaton. Electromyographic control of a hands-free electrolarynx using neck strap muscles. *Journal of Communication Disorders*, 42:211–25, 2009.

[20] Eivind Kvedalen. Signal processing using the Teager energy operator and other nonlinear operators. Master's. thesis, University of Oslo Department of Informatics, 2003.

[21] Sami Lemmetty. Review of speech synthesis technology. Master's thesis, Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Finland, 1999.

[22] R. McGill, J.W. Tukey, and W. Larsen. Variations of box plots. *The American Statistican*, 32:12–16, Feb. 1978.

[23] Keigo Nakamura, Matthias Janke, Michael Wand, and Tanja Schultz. Estimation of fundamental frequency from surface EMG. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 573–576, 2011.

[24] WebMD Network. Medterms dictionary. *http://www.medicinenet.com*, 2013.

[25] Frank Neudecker. The influence of signal smoothing filters on brake test accuracy. Technical report, Corrsys Datron, 2004.

[26] Katsutoshi Ooel, Carlos Rafael Tercero Villagran, and Toshio Fukud. Development of the compact control system using of neck EMG signal for welfare applications. In *Proceedings of International Symposium on Micro-NanoMechatronics and Human Science (MHS)*, pages 127–132, 2010.

[27] Petr Pollák and Josef Rajnoha. Long recording segmentation based on simple power voice activity detection with adaptive threshold and post-processing. *International Conference on Speech and Computer, St. Petersburg*, June 2009.

[28] M.B.I. Reaz, M.S. Hussain, and F. Mohd-Yasin. Techniques of EMG signal analysis - detection, processing, classification and applications. *Biological Procedures Online*, March 2006.

[29] Cara E. Stepp, James T. Heaton, Rebecca G. Rolland, and Robert E. Hillman. Neck and face surface electromyography for prosthetic voice control after total laryngectomy. *IEEE transactions on Neural Systems and Rehabilitation Engineering*, 17:146–55, April 2009.

[30] Bernard Widrow and S. D. Stearns. *Adaptive Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ, 1985.