# Myoelectric Signal Processing for Prosthesis Control

**Dissertation**

zur Erlangung des mathematisch-naturwissenschaftlichen
Doktorgrades
"Doctor rerum naturalium"
der Georg-August-Universität Göttingen

vorgelegt von

**David Hofmann**

aus Vipiteno, Italien

Göttingen 2013

**Prof. Dr. Florentin Wörgötter (Referent)**
Drittes Physikalisches Institut, Georg-August Universität Göttingen

**Prof. Dr. Dario Farina (Referent)**
Department of Neurorehabilitation Engineering, University Medical Center Göttingen

**Prof. Dr. Marc Timme**
Network Dynamics Group, Max Planck Institute for Dynamics and Self-Organization, Göttingen

# Acknowledgements

I thank J. Michael Herrmann for giving me the possibility to work at this great institute and supervising my work. I am grateful for your support that reliably grew exponentially as deadlines were approaching and for many interesting and stimulating intellectual conversations beyond the actual work.

I would like to thank Theo Geisel for his effort to create such an excellent and creative environment at his department for Nonlinear Dynamics. There is no doubt: This is a very special place.

I am grateful to my thesis committee, Florentin Wörgötter, Dario Farina and Marc Timme for guiding my work and sincere support whenever I needed it. Additionally, I would like to thank Hansjörg Scherberger for joining my thesis examination board.

I thank also Fred Wolf who gave me the opportunity to work on another interesting topic besides machine learning and hand prosthesis. You taught me very valuable principles about the scientific method in those few but intense interactions. At this point I would also like to thank Andreas Neef who proved to be an invaluable supervisor and colleague at the same time.

I thank Yorck Beensen for his valuable support for just about every computer-related issue I had. I fear no administrator I will encounter in my future path will ever reach your level. At this point I would also like to thank Denny Fliegner who was always forthcoming in answering questions and who keeps the computing equipment at such a high standard clearly facilitating each ones work.

Many thanks to Ayse Bolik who accomplishes virtually every bureaucratic issue in less than a minute - your competence and helpfulness are invaluable. Additional thanks to you and also Regina Wunderlich and Viktoria Novak for your great support with every organizational efforts, especially the organization of the summer school for computational neuroscience.

I very much want to thank Mirko for being a great office mate and friend throughout the years! Uncountably many times you kept my mood up, you stimulated interesting political discussions and it is you who kept my Italian alive in Göttingen. I could not have had more luck than this, sharing with you the entire - busy and intense - PhD time.

I also very much thank Wolfgang who started off being my Jazz piano hero, later became my climbing partner - many moments we shared at the wall, many moments of mutual trust and dedication to overcome our own limits - and invaluable friend. Göttingen was better with you!

Finally I thank my mother for her constant support throughout my life and her brave heart that lets her only son go wherever his way takes him.

# Contents

# 1 Introduction

Understanding the control of the complex human biomechanical machine is a great challenge; not only from an academic and intellectual perspective, but also because of its potential to provide insights which may help clinicians to improve the quality of treatment of motor disorders. Besides the application of myoelectric signals for the control of prosthesis which we pursue in this work and will introduce in section §1.3, there is a manifold of other medical and also non-medical applications.

Recently, advanced multifunction upper limb prostheses received a lot of public media attention often hyped as "the mind-controlled artificial hands" [3]. Advanced signal processing of the electromyographic signal and innovative surgical procedures, such as the targeted muscle reinnervation, are the driving forces behind these achievements. However, the idea of myoelectric prostheses is more than half a century old and, nonetheless, and despite the enthusiasm of some public presentations of artificial devices, there still exist considerable challenges before these developments can be beneficial to the general amputee population [83].

and extend an advanced approach to amplitude estimation which is a central feature used extensively in clinical and non-clinical applications of myoelectric signal processing. Furthermore, we study several issues related to the pattern recognition approach to myoelectric prostheses control. We investigate spatial filters to alleviate two central problems of the standard pattern classification approach: the performance-delay trade-off and the non-stationarity of the signal during dynamic contractions. Furthermore, we study algorithms for electrode selection in order to address two questions: how many electrodes can be added, and still provide an improvement in control accuracy and where should these channels be positioned to obtain good control accuracies.

## 1.1 Historical notes on electromyography

The pioneers of electromyography are known to be Luigi (Aloysius) Galvani "De Viribus Electricitatis in Motu Musculari Commentarius" (1792), Alexander von Humboldt "Versuche über die gereizte Muskel und Nervenfaser" (1797) and Guillaume Duchenne (de Boulogne) "De l'électrolisation localisée" (1872). Today, Galvani is considered the oldest source in electromyography but original documents indicate that contemporary scientists were working on the problem of muscular function already in the late 17th century

(see [35] for a systematic review[1]).

In 1849, six decades after Galvani had documented the finding that electric currents could initiate muscle contractions, DuBois-Reymond discovered that it was also possible to record electrical activity during a voluntary muscle contraction. DuBois-Reymond's achievement was an example of scientific intransigence. He devised a surface electrode which consisted of a wire attached to a paper immersed in a jar of saline solution. When the fingers were immersed in the saline solution and the arms and hand were contracted, the deflection on the galvanometer was minute (approximately 2° to 3°). Realizing that the impedance of the skin reduces the current that drives the galvanometer, he induced blisters in each forearm, removed the skin and placed the open wounds in contact with the saline solution of the electrode. Upon contraction he now measured a sizable deflection (65°) on his galvanometer. He repeated the contraction three times for each arm and always obtained a similar result. To remove doubt, he repeated the whole experiment several weeks later, after the original wounds had healed and obtained again the same results. [4]

Later, in 1890, Marey made the first recording of electrical muscle activity and introduced the term *electromyography* [35]. In 1922, Gasser and Erlanger used a cathode ray oscilloscope in place of a galvanometer, which was up to then the standard method to show the electrical signals from muscles. This feat won them the Nobel Prize in 1944 [38].

However, because of the stochastic nature of the myoelectric signal only rough information could be obtained from its observation at that early days mainly due to the insufficiency of the measurement apparatus. Only in the 1960s the quality and availability of electronics apparatus eventually reached a level that enabled anatomists, kinesiologists, and orthopedic surgeons to make increasing use of electromyography.[38]

## 1.2 Myoelectric signals and their use in medicine

### 1.2.1 Myoelectric signals

Myoelectric signals are a side effect of muscle activity. When active, muscle cell's membrane potential changes similar to neurons when they generate an action potential. Hence, the membrane potential change stems from a change in intra-cellular vs. extra-cellular ion concentration mediated by ion currents driven by a gradient of the electrochemical potential [116, 57]. The ions pass the cell membrane through voltage-gated

---

[1]We wish to note that there are some inconsistencies among the relevant literature not resolved in [35] concerning the contribution by Francesco Redi, who was perhaps the first scientist to document electricity produced by muscles. His studies of electric ray-fish made him deduce that the specialized electric organ was essentially a modified tissue of ordinary muscle [135, pp. 47-51]. However, different authors [4, 133] cite an apparent document of F. Redi from 1666 which according to [35] is not to be found. The document referring to the aforementioned finding is from 1671 [135] as cited in [157, 38] and is freely available from Google Books.

ion channels resulting in a membrane voltage change. This in turn leads to additional openings of channels and, hence, is a self-reinforcing process which is followed by self-inhibition mediated by voltage dependent inactivation of these ion channels. On the whole, this mechanism produces the well known all-or-none "firing" event [86].

The depolarization that initiates this event is generated by the $\alpha$-motor neuron that controls the activity of that muscle cell and is initiated by the influx of calcium ions at the so called motor end-plate which is a big chemical synapse of the muscle cell. The $\alpha$-motor neuron together with all muscle cells that it innervates form a small functional group called a motor unit (MU). Each muscle fiber is normally innervated by only one motor neuron in only one place, usually near its midpoint. The motor end-plate carries enough vesicles filled with the neurotransmitter acetylcholine to assure a reliable[2] action potential (AP) transmission from neuron to muscle fibers. Under normal conditions, a motor neuron AP elicits an AP in each of its innervated muscle fibers. Hence, muscle fibers belonging to the same unit fire synchronously and the resulting action potential waveform is usually called a motor unit action potential (MUAP). Their fibers are scattered over a substantial part of the muscle and the extracellular action potential shape of a muscle cell is biphasic so the resulting shape of a muscular unit action potential will necessarily dependent on the place of the measurement site.

Motor neurons have their cell body located in the spinal cord and all motor neurons innervating the same muscle form a motor nucleus which is extending over one to four spinal cord segments[3]. As motor units are of different sizes (larger motor neurons innervate also more muscle cells) and given that larger $\alpha$-motor neurons need a higher total current to make them fire, the order of recruitment of motor units is naturally set and normally called the Henneman's size principle. [86]

The central nervous system controls the force generated by a muscle through the number of recruited MUs and the modulation of the firing rate of the motor neuron pool [48]. A single AP of a motor neuron results in a small twitch of the muscle cells it innervates. If the firing rate increases, the twitches will start to merge into each other and the average force produced by the motor unit will steadily increase. However, as the firing rate is progressively increased, the twitches will get closer to tetanus, at which point no further force can be produced by the muscle. Hence, the relationship between stimulation frequency and force for single motor units is nonlinear [19].

We have seen that muscle cells belonging to the same MU fire synchronously. Synchronicity of several MUs in turn depends on the neural control of motor neuron pools as well as feedback from muscle spindles [41]. Muscle spindles are sensory neurons in the muscle that project on the $\alpha$-motor neurons and change their firing rate according to external load on the respective muscles. This way a homeostatic feedback circuit is formed between spine and muscle to stabilize contractions. A certain degree of synchronicity of

---

[2]This is in contrast to chemical synapses in the cortex where synapses are known to be unreliable and, hence, often seen as major noise source [86]

[3]The human spinal cord is made up of 31 spinal segments which are defined by bilateral pairs of nerve roots (note that they are not equivalent to vertebral segments) [86].

| Property | Numbers |
|---|---|
| Number of muscle fibers within the brachioradialis | $> 129200$ [48] |
| Number of $\alpha$-motor neurons innervating the brachioradialis | $315 - 350$ [48] |
| Number of muscle fibers innervated by an $\alpha$-motor neuron | $25 - 2000$ [48] |
| Size of a muscle fiber (diameter/length) | $50 - 100 \ \mu m/2 - 6$ cm [86] |
| Muscle cell action potential propagation velocity | $3 - 5$ m/s [86] |
| Firing rate of $\alpha$-motor neurons | $5 - 30$ Hz [41] |

Table 1.1: Some numbers about muscles. These are meant to give an idea about the order of quantities. Naturally, number of muscle fibers and $\alpha$-motor neuron are highly dependent on the volume of the muscle (densities of these quantities could not be found in the literature research) therefore numbers for a specific lower arm muscle, the Brachioradialis, is given.

MUs results in larger force levels while too synchronous firing gives rise to tremor [41]. However, the synchronous behavior of the many MUs results in small (in the microvolts) but still measurable electrical signal at the skin surface, i.e. the electromyogram (EMG) and, as we may conclude, is correlated to some degree [51] with the muscle force.

## 1.2.2 Measuring the myoelectric signals

We can consider the skin as the boundary between two media, a conductive layered non-homogeneous and anisotropic semispace (skin, subcutaneous tissue, and muscle) that contains sources of electric field, and an insulating semispace (air). The sources of electric field (ion currents) generate a two-dimensional potential distribution on the surface of the skin. The potential is defined with respect to a reference point at a sufficient distance from the sources so that it is not influenced by them. The ideal condition for measuring this potential distribution is to have a point electrode connected to an infinite input impedance voltmeter measuring the voltage with respect to a remote reference where the potential is zero.

The departure from the idealized scenario is manifold: The electrode is a two dimensional metal object that forces the area of contact to be equipotential and therefore modifies the skin potential distribution in its neighborhood, the skin-electrode contact has a frequency and current dependent impedance, and the voltmeter (amplifier) has a finite input impedance. In addition, other sources of potential contribute to the reading. Among these are DC and noise voltages generated at the skin electrode interface, the capacitively coupled power line voltage, cable motion artifacts and other electrical phenomena unrelated to EMG. [108]
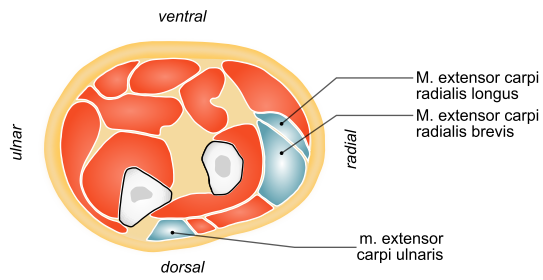
(a) Extension
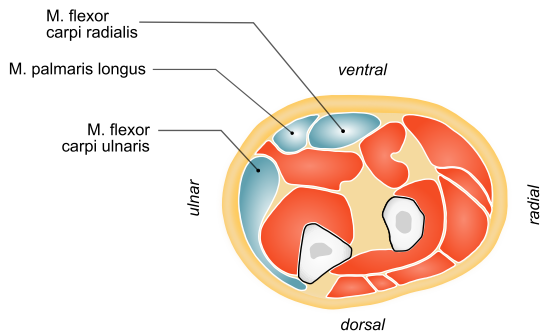


(b) Flexion
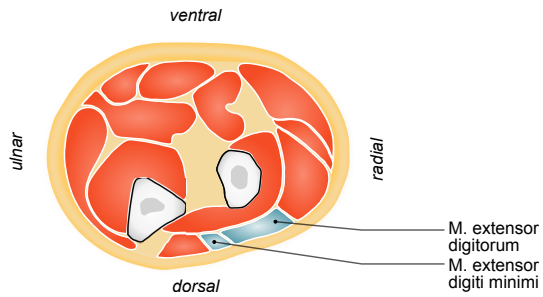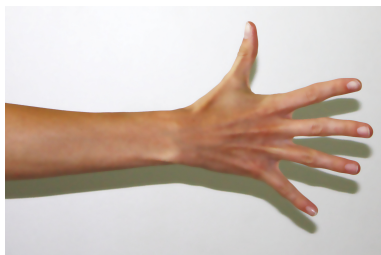


(c) Hand open
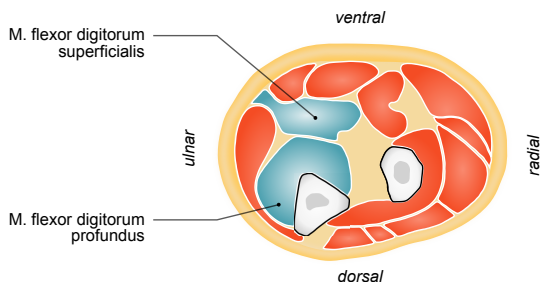


(d) Hand close

Figure 1.1: Hand and wrist movement with its responsible muscles. Cross section at one third (proximal) of the lower limb. (adapted from [137])
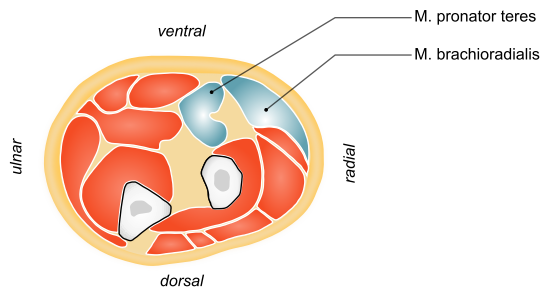
(e) Pronation
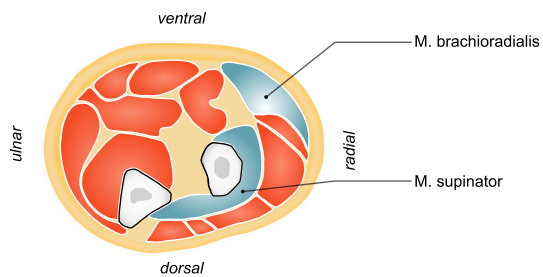


(f) Supination
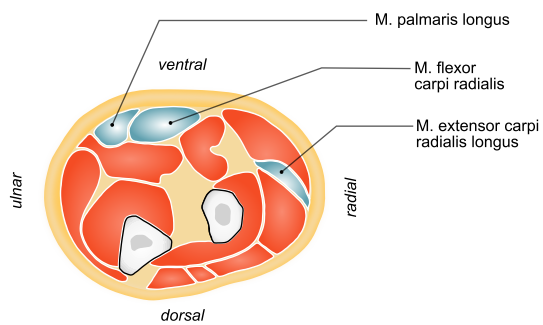


(g) Abduction
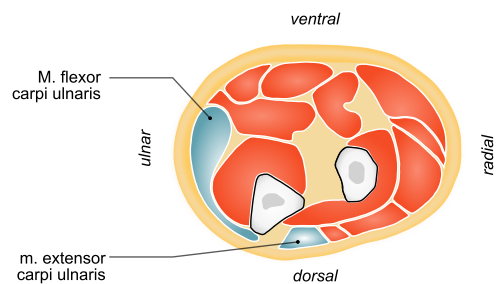


(h) Adduction

Figure 1.0: Hand and wrist movement with its responsible muscles. Cross section at one third (proximal) of the lower limb. (adapted from [137])

In order to reduce or remove some noise sources, the following provisions can be taken. Active electrodes, i.e. electrodes with a differential amplifier, are recommended as they almost completely eliminate cable motion artifacts [34]. The power-line noise is removed by a band stop filter with the respective frequency. Additionally the signal is band-pass filtered with cutoff frequencies at 5-20 Hz and 500 Hz to extract the frequency range which is seen to be the relevant band for EMG [108]. Silver or silver-chloride electrodes are suggested as when in contact with skin they are known to form an almost resistive impedance in the EMG frequency range, while other metals present capacitive components that introduce additional filtering [108]. Conductive gel may be applied in order to reduce the impedance between skin and electrode [34]. After attenuation of the unwanted motion artifacts and power-line interference, the signal is treated as though it were free of artifacts, although an additive background noise remains [108, 34].

### 1.2.3 Myoelectric signals for medical applications

The topic of this thesis is powered prostheses control using myoelectric signals. This is one important medical application of myoelectric signals which will be discussed in section §1.3 and with the focus on pattern recognition approaches in chapter 3 of this thesis. A strongly related topic to prostheses control is orthoses control. While a prosthesis replaces a body part, an orthosis is an artificial device that aids or supports the movement of a weak or injured body part. Orthosis control efforts are often approached with neuromuscular modeling. However, neuromuscular models can be high dimensional and ill-posed [93]. Then measurements with motion capturing techniques do not suffice for obtaining reliable parameter values and it follows that electromyographic measurements as estimates of muscle forces are essential for the purpose of neuromuscular modeling [93]. These models then are employed for rehabilitation purposes, for instance by the analysis of walking patterns [82] or by controlling robots (ex. powered orthoses) whose purpose it is to assist treatments in physiotherapy [140].

Myoelectric signals are also used in the diagnosis of chronic pain (ex. muscle induced headache or back pain) [39]. Besides this so called trigger points are investigated with EMG. These trigger points, when active, are known to be responsible for muscle pain. EMG measurements can help locate them (as spontaneous muscle activity is increased in an active trigger point) and hence facilitate their treatment [37].

Furthermore, it is known from myoelectric studies that psychological disease, as for example depression, have as a psychosomatic symptom an elevated muscle tension [37]. Besides trying to quantitatively classify psychosomatic disease through myoelectric measurements, these findings also lead to the attempt to treat depression with the aid of acoustic and other forms of biofeedback based on EMG (see [155] as cited in [37]). And finally, myoelectric signal (MES) are also employed for the study of Parkinson disease as described in [151].

All these applications make use of the myoelectric signal amplitude estimates. We devote chapter 2 to the study and comparison of techniques for amplitude estimation

MES hoping that the results might be beneficial not only to prosthesis control but also to other applications.

### 1.2.4 Further applications of myoelectric signals

Additional examples of application were one wishes reliable control based on myoelectric signals are briefly discussed in what follows. In 1990, Knapp and Lusted designed a computer system to acquire myoelectric signals, perform feature extraction on these signals, and then map the desired features to MIDI commands in real-time. They called this system Biomuse and state that it can be used as a "biocontroller" to augment normal musical instrument performance or as a direct human-computer interface for musical composition and performance [94].

More recently, virtual environments get more attention and MES are among those signals taken into consideration for controlling objects in the virtual environments [139]. By advances that allow the detection of single finger movements from myoelectric signals telemanipulation [134, 139] becomes a further possible application or even the control of computer games [120]. Finally, scientists have shown that MES can be used for augmenting speech recognition [20].

## 1.3 Myoelectric prosthesis control

Its history starts with Reinhold Reiter in Germany in 1945 who applied for a patent describing the concept to employ electromyographic signals for controlling prostheses [4]. This was still at the time where amplifiers where based on cathode ray tubes and the silver-silver chloride electrode was not available. Indeed it was only in the 1960s, when a group of Russian engineers led by Kobrinsky revealed the design of a hand prosthesis controlled by myoelectric signals detected from the forearm muscles [4]. This demonstration excited the engineering and rehabilitation community and hence became the first landmark in the field.

However, despite the many efforts taken since this landmark in myoelectric prosthesis control, the current state of the art only scrapes through the criteria for clinical relevance. Already in 1973 Childress states in a review article on clinical significance of powered limb prostheses [23]:

> "In its fullest sense the clinical use of a limb prosthesis means that it can be prescribed for a patient, purchased for or by the patient, and fitted to the patient; and that training and maintenance may be provided in its use. A prosthesis is not clinically significant until all these requirements may be satisfied. The clinic is the final testing area for a prosthesis. If it is not usable by patients, it is of little value no matter how technologically marvelous or ingenious it may be."
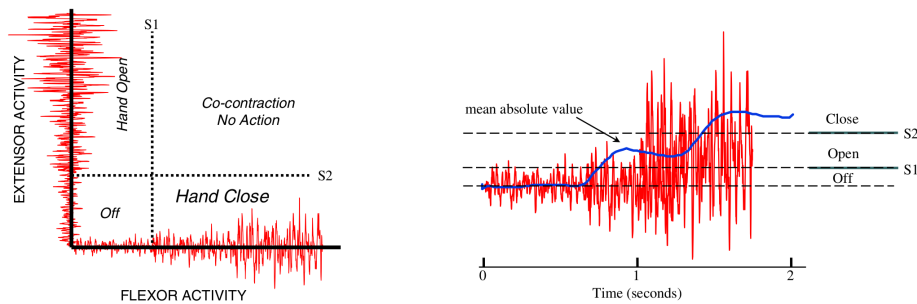
Figure 1.1: Conventional control schemes (adapted from [108]). **Left:** two-state amplitude controller. **Right:** three-state amplitude controller.

Even though conventional myoelectric prostheses meet these criteria, its functionality is not high and many amputees thus prefer a cosmetic non-powered prosthesis over a MES controlled powered prostheses as limb replacement. Indeed, whether or not an artificial limb is an acceptable replacement for a human limb depends especially upon the expectations of the affected individual, their motivation to incorporate this device into their lifestyle, and the functionality of the device. Clearly, these issues are not mutually exclusive: a potential user will be more motivated to learn how to use a highly functional device controllable in an intuitive way.

In conventional myoelectric control schemes a function is activated by comparing the EMG amplitude to a threshold and different functions are controlled by the same technique applied to multiple recording sites. Two control schemes based on this threshold approach are employed for prosthesis control (see figure 1.1). Both of these approaches are inherently limited. A two-state amplitude controller needs two independent signal sites per degree of freedom. Depending on the level of amputation there might simply not be many muscles and hence signal sites available. And even in the most convenient case where all arm muscles are still intact it is clear that hand and wrist contractions are controlled by muscles in a synergistic rather than independent way (confront figure 1.1). Probably an amputee could learn to use the muscles independently but this would result in a lengthy physiotherapeutical training period. Another approach besides using more independent signal sites is to switch sequentially through degrees of freedom to be controlled by the same two signal sites by crossing the two thresholds S1 and S2 simultaneously (co-contraction regime in the left scheme in figure 1.1). This way one degree of freedom can be controlled at a time, however, it is clear that increasing the number of states quickly makes the control of the prosthesis cumbersome.

To reduce the number of channels needed per degree of freedom from two to one the three-state amplitude controller was proposed. For the threshold approach this is the limit of controllable functions possible per channel [125]. However, even if the three-state controller increases the functionality it has the drawback being non-intuitive because a contraction level above the first threshold will provoke the controller to exert a certain

function and to achieve the antagonistic function the contraction force level must be increased in order to cross the second threshold S2. This is a controlling style for which the amputee needs to do additional training which will become tedious if multiple degrees of freedom should be controlled simultaneously.

With the advent of machine learning algorithms, most notably the perceptron in 1958 [136], an alternative approach became slowly appealing to the community of myoelectric prosthesis control. Namely the pattern recognition or machine learning approach which is intimately related to the mathematics field known as statistical learning theory [60]. This was and still is seen a promising approach that is not suffering from the drawbacks of the amplitude threshold controller and, hence, might lead to increased functionality and an intuitive control. However, yet there is no myoelectric prosthesis based on a pat-



Figure 1.2: Otto Bock Michelangelo hand, providing four degrees of freedom of the wrist and hand[83].

tern recognition system that passed clinical tests. This stands very much in contrast to the apparent academic achievements. Interestingly, myoelectric controllers based on a big variety of pattern classification systems have appeared in the literature since the 1990s and have provided high performances ($>$90% classification accuracy) in off-line and on-line tests. This dichotomy between academic research and clinical application may require a general paradigm shift as proposed by Jiang, et al. [83]. The authors advocate the necessity of closing the loop, i.e. developing feedback systems from the artificial limb to its user for instance by cutaneous electrical stimulation, and by extending sensor types beyond myoelectric electrodes, hence, develop multi-modal prostheses. These are long term projects that can certainly benefit from the further development of the pattern recognition approach (whether it being classification or regression approaches).

Let us turn back to the more immediate issues faced with pattern recognition approaches that inhibit the transfer of apparently high performance in academic research to clinical application. The problems are easily understood in view of the hypothesis the pattern recognition approach relies on, namely the myoelectric signals contain patterns which are specific for muscle contractions responsible for the respective intentional limb postures or motions. The major drawback is the dependency of classification performance on what is generally subsumed with the termed "signal non-stationarities". Indeed, pattern recognition systems after training, i.e. modeling the statistics of the signals, usually demands stationarity of the signal distribution. However, several influences listed in the following cause changes of these statistics, i.e. signal patterns, and lead thus to a significant performance deficiency. The most important issues are recognized to be:
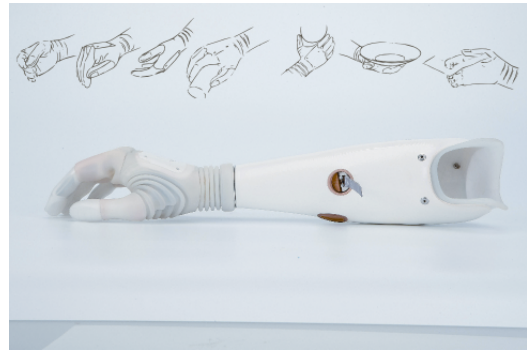
- transient regime between static contractions [103, 71]

- electrode shifts [68]

- impedance changes due to changes in skin conditions (e.g. sweat)

- muscle fatigue

- upper arm position [58]

The latter is important for transradial prostheses because the position of the upper arm affects the control signals at the lower arm, hence, patterns for the same hand contraction may be different, because for the control of an arm posture, weight must be borne against gravity. Therefore, muscle activation necessarily varies for different positions even though the same contraction is performed.

A general and straight-forward way to deal with these signal non-stationarities is to measure signals for the respective condition and include them into the training data of the pattern recognition system in the hope that the learning algorithm adapts to these deviations from the patterns in the canonical case [68, 103, 71]. While with canonical case we refer to the EMG during an isotonic contraction which is used by most studies. Indeed this seems to be the major cause for the dichotomy between high performances in terms of classification accuracy and the fact that yet there is no prosthesis based on pattern recognition systems in clinical use.

## 1.4 Thesis overview

Myoelectric signals are stochastic in nature and, hence, statistical methods appear as the appropriate tools for the processing of these signals. In this way it becomes possible to extract the features that are indicative for clinical studies or relevant for the control of electric devices such as prostheses. In this thesis we, therefore, start with probabilistic modeling of myoelectric signals in section §2.2. We introduce a mixture distribution that is interesting for an analytical formulation of a crucial step of the signal processing algorithm. We then proceed in section §2.3 with the first of three main contributions of this thesis, namely, an advanced amplitude estimation technique termed Bayes-Fokker-Planck filter. After introducing the general pattern recognition approach to prosthesis control, we focus on filtering and feature selection which form two important sub-problems within that framework. More specifically, these are the analysis and comparison of spatial filters with respect to the improvement of the classification accuracy in section §3.3 and feature selection methods for the purpose of electrode selection in section §3.4. We conclude by summarizing and discussing our findings in chapter 4.

# 2 Amplitude estimation of myoelectric signals

We have seen that myoelectric signals are fast varying signals even when the muscle tension does not vary. Hence, if not preprocessed, they are not suited as a control signal for any device that needs to be controlled in a deterministic fashion as, for example, an artificial limb.

A suggestive way to smoothen a fast changing signal is to apply a low-pass filter. Indeed, what the large majority[80, 97, 98, 72, 73, 101, 55, 123, 42, 32, 143, 26, 29, 28, 34, 52, 99, 130, 87, 88] of studies dealing with amplitude estimation do is exactly this. The hope is that the resulting signal can be used for reliable control. Reliability is usually quantified by signal-to-noise ratio (SNR) of the amplitude estimate. This gives an index of how discriminable different amplitude levels are which correspond to different contraction forces.

Another branch of studies tightly related to amplitude estimation are those dealing with force or torque estimation from the myoelectric signal [123, 42, 129, 146, 145, 138, 144, 147]. Here, quality of the estimate is given in terms of root mean square error (RMSE) [138], correlation coefficient $r$ [129] or coefficient of variation $r^2$ [138] as a target value, namely force or torque, is measured. Indeed, EMG amplitude estimation is the most common experimental means to estimate muscle forces and the neural drive to a muscle (see [80, 109, 153, 141] as referenced by [88]), despite its known unreliability, and eloquently stated [42]:

> "One of the most frustrating, or appealing (depending on your perspective), aspects of the surface EMG signal is that when rectified and sufficiently smoothed, its amplitude is qualitatively related to the amount of torque (or force) measured about a joint, but more often than not, an accurate quantitative relationship is elusive."

Also as an estimate for the neural drive to a muscle the amplitude seems a poor measure due to amplitude cancellation as reviewed in [51]. Nevertheless, it is successfully used as a control signal as will be discussed in this thesis.

In the following we will overview the most important studies that treat the temporal filtering problem.

## 2.1 Literature overview

In myoelectric signal processing most approaches to amplitude estimation are synonymous to standard deviation estimation and differ mostly in the kinds of preprocessing and number of electrodes used. The generally accepted (see table 2.1) phenomenological model on which those procedures are based is formally stated as

$$s(t) = \sigma(F(t)) \cdot (H * \eta)(t). \tag{2.1}$$

The idea is that the myoelectric signal $s(t)$ depends on $\sigma(F(t))$ which is the force dependent amplitude that is to be estimated [72]. Technically this amplitude is the scale of a stochastic process $\eta(t)$ that is filtered by a time independent filter $H$. Let us briefly discuss assumptions about those quantities:

- $\eta(t)$ is an uncorrelated standard normal distributed stochastic process $\eta(t) \sim \mathcal{N}(0, 1)$, $\langle \eta(t + \tau)\eta(t) \rangle = \delta(\tau)$ [72, 34]. The assumption of Gaussianity of the process is justified by the application of the central limit theorem to the superposition of many (see table 1.1) muscle fiber action potentials [72]. In 1999, however, Clancy and Hogan have shown that "the experimentally observed densities fall in between the theoretic Gaussian or Laplacian densities" [33]. Hence, the functional form of the probability density from which $\eta(t)$ is drawn is still unclear. We will analyze and discuss the probability distribution of myoelectric signals in detail in chapter 2.2. Note that in our discussion as well as in [33] the signal is assumed to be stationary, i.e. isotonic contractions are performed.

- $H$ is a time independent filter function accounting for the effects of tissue layers and electrode properties as well as intrinsic firing rate limitations of MUs. Filtering limits the frequency band of the stochastic process and thus affects the estimation error [11]. To avoid this and, hence, improve the amplitude estimation, the signal is whitened by estimating the inverse of the filter function, $H^{-1}$, from data. This requires a calibration step that may sometimes be impractical but leads to improvements of $\sigma$ estimation in terms of SNR [72, 30, 31].

- $\sigma(F(t))$, the scale of the stochastic process, is a time homogeneous (memoryless) function depending on the force $F(t)$. The dependence on $F(t)$ is usually modeled by some nonlinear function [72, 55, 19] but the efforts to quantify this relationship are not reliable [42]. Some authors [138, 72] interpret $\sigma$ as the neural drive to the muscle. More recently it was shown (see [87, 40, 51]) that due to amplitude cancellation the amplitude estimation necessarily deviates from the neural drive as given by the extraction of motor unit action potential trains (MUAPTs). Indeed amplitude cancellation is at least in part responsible for the infeasibility of force and torque estimation [87, 88, 51]. However, these facts still allow the estimated $\sigma$ to be used as a control signal for proportional and classifier based

control as discussed below. But before we come to that, we turn our discussion towards the statistical estimation problem and review some alternative approaches to amplitude estimation.

Given this phenomenological model probably the most straight forward approach to estimate $\sigma$ is to deconvolve the signal $s$ with the inverse filter $H^{-1}$ and then estimate the variable of interest by for example the maximum likelihood estimator. For the standard deviation of a Gaussian process this is given by the ensemble average of $N$ repetitions of the stochastic process

$$\hat{\sigma}(t) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i(t) - \mu(t))^2}. \tag{2.2}$$

$x_i(t)$ denotes the $N$ times repeated stochastic process, $\hat{\sigma}(t)$ is the estimation of the true $\sigma(t)$ (for the sake of simplicity we omit from now on the dependence on $F(t)$) and $\mu(t)$ is the mean of the process. In the case of myoelectric signals $\mu(t) = 0$ [72].

Indeed, an identical repetition in terms of probability distribution of the myoelectric signal of dynamic contractions is unlikely to be achieved or at least very difficult and in any case very tedious as $N$ should be large for a good estimation. However, during an isotonic contraction we assume the process to be stationary [31, 28, 55, 72]. Thus we may exchange the ensemble average in (2.2) by a time average and avoid the $N$ repetitions. Further, we discretize time given that the measurement procedure samples the signal at some sampling rate $r$. We get

$$\hat{\sigma} = \sqrt{\frac{1}{T \cdot r} \sum_{i=1}^{T \cdot r} s_i^2} \tag{2.3}$$

where $T$ denotes the time window length for which the estimation is computed. This is the root mean square (RMS) of the signal and is only an estimator of the standard deviation if above considerations of zero mean and stationarity are valid. Importantly, for short enough time windows $T$ stationarity of the signal is often assumed and hence the estimator (2.3) will give a good approximation to $\sigma$ even for non-isotonic contractions.

From this equation we can learn two aspects of amplitude estimation:

1. Estimation of $\sigma$ is equivalent to temporal filtering the process with a degree equal to $T \cdot r$.

2. The quality (in terms of precision, i.e. inverse of variance) of our estimation will depend on the time window $T$ and on the sampling rate $r$.

Thus for nonstationary signals a trade-off between estimation variance ($T$ short) and violation of stationarity assumption ($T$ long) will limit our estimation accuracy and put

constraints on the choice of $T$. The time window length was investigated for instance by [26] and resulted in an adaptive filter. However, according to simulation studies by Keenan and Valero-Cuevas [88], the necessary time window length for a good estimation is additionally increased by amplitude cancellations due to the summation of MUAP waveforms.

A further, quite interesting finding was presented in 2004 by Potvin and Brown [129]: force estimation is substantially improved if up to 99% of EMG signal power is removed before estimating the amplitude. The authors use a high pass filter with cut-off frequencies of up to 440 Hz which very much contrasts the widely established view that the relevant frequency window lies between 20 and 500 Hz. In support of this finding, Staudenmann et al. [145] demonstrate that principal components with intermediate, not those with the largest eigenvalues give the best results for force estimation.

Potvin and Brown point out [129] that high pass filtering is related to whitening because both procedures flatten the power spectrum. Indeed, the major part of myoelectric signal's power lies between 20 and 500 Hz with peaks around 70 to 100 Hz and decaying towards 500 Hz [100]. Hence, all peaks in the power spectrum are attenuated by a high-pass filter leaving only 1% of the original signal power. Thus also high pass filtering flattens the power spectrum and removes temporal correlations to a certain extent.

Another explanation for why high pass filtering improves force estimation is that amplitude cancellations are more dominant for low frequencies as stated by Keenan and Valero-Cuevas [88].

### 2.1.1 Methods for nonstationary signal analysis

As discussed earlier the classical approach for amplitude estimation is in general not suited for nonstationary signals. Clancy proposed an adaptive filtering procedure to dynamically adapt the time window length which improves the estimation [29]. Apart from this, two studies applied a Kalman filter to myoelectric signals [64, 49]. Kalman filtering is a method for nonstationary signal tracking that we will discuss in more detail in chapter 2.3. However, we wish to point out two aspects of those approaches. In the study by Graupe and Cline[64] even though a Kalman filter was used it was put on top of an auto-regressive moving average (ARMA) model which demands stationarity of the signal. Hence, the proposed method is properly applicable only to stationary signals. The most recent approach to nonstationary MES amplitude estimation was presented by Terrence Sanger in 2007 [138]. He proposed a specific Bayesian filter that we are going to study in detail in 2.3 and apply to our data. He compared the Bayes filter to the optimal linear filter as well as low-pass filters with cut-off frequencies of 0.1, 1 and 5 Hz and it turned out that the Bayes filter with a Laplace likelihood performed 3 times better, in terms of SNR, than the 1 Hz low-pass filter which was the best among the linear filters tested. The method is thus ideally suited for the purpose of myoelectric control and discrimination of different force levels. However, before we discuss this algorithm we investigate the probability distribution of MES in detail.

| publication | sEMG model | estimation approach | conditions (static/dynamic) |
|---|---|---|---|
| Graupe, et al. [64] | ARMA, additive noise | ARMA and Kalman filter | static |
| Hogan and Mann [72, 73] | eq. (2.1), noise ignored | low pass filter, after spatio-temporal whitening | static |
| Evans, et al. [49] | eq. (2.1), noise ignored | Kalman filter, after log-transform | dynamic |
| Clancy and Hogan [31] | eq. (2.1), noise ignored | low pass filter, after spatio-temporal whitening | static |
| Clancy and Farry [29] | eq. (2.1), with additive noise | low pass filter, after temporal and adaptive whitening | static |
| Clancy, et al. [28] | eq. (2.1), with additive noise | low pass filter, after spatio-temporal whitening | dynamic |
| Potvin and Brown [129] | not discussed | band pass filter: high-pass cutoff above 250Hz | static |
| Staudenmann, et al. [145] | not discussed | low pass filter, PCA for spatial processing | static |
| Sanger [138] | stochastic variable; compares following prob. dist: Gauss, Laplace, Poisson | Bayesian filter with Fokker-Planck evolution | static |

Table 2.1: Amplitude estimation publication overview.

## 2.2 Statistical properties of myoelectric signals

Myoelectric signals are in general non-stationary[64] noisy signals. Noise sources are intrinsic (different tissue layers between muscle and electrode) and extrinsic (measurement device). Some authors claim noise to be additive[64], while others argue it to be multiplicative[98, 72, 49]. Given the assumption that myoelectric signals are a sum of action potentials from many independent[1] muscle cells one would naively conclude that the resulting distribution is Gaussian due to the central limit theorem. This is what frequently is being assumed. However, different evidence in the literature does not allow a decisive conclusion on this matter yet, as we will review next.

Milner-Brown and Stein [109] reported that the distribution of the MES from isotonic contractions of the first dorsal interosseus muscle was more sharply peaked near zero than a Gaussian distribution. Recordings at higher force levels tended to appear less peaked than those at lower force levels. Parker et al. [124], using fine wire electrodes inserted into biceps muscles, graphically compared the MES probability density to a Gaussian density during light and moderate contraction levels. They concluded that the MES is reasonably modeled as a Gaussian random process. Hunter et al. [78] using surface electrodes on the biceps muscles, graphically compared the MES probability density to a Gaussian density. Isotonic contractions were conducted at 30% MVC. They found that the density function departed considerably from the shape of a Gaussian, being more sharply peaked near zero. Bilodeau et al. [12] examined contractions of the biceps muscles. Both constant-force (20%, 40%, 60%, 80% MVC) and slowly-force-varying contractions were studied. Using a Shapiro-Wilk test, they generally found that MES present a non-Gaussian distribution, being more peaked than a Gaussian distribution. Higher force levels tended to give rise to slightly more Gaussian distributed signals. Finally, Clancy and Hogan [33] propose the symmetric Laplace distribution as an alternative model to the Gaussian. They conclude that the MES distribution lies "in between" the two proposed models.

In the following chapter we study the measured probability distribution in detail and point out that not only the peak of the distribution departs from that of a Gaussian but also its tails and additionally MES often have a strong skewness from which we conclude that neither a Gaussian nor the symmetric Laplace distribution as studied in [33] can be an adequate model given their symmetry.

### 2.2.1 Probability density function

The probability density function of a continuous stochastic variable fully characterizes its properties. If it is given we may derive Bayes optimal classifiers, information theoretic properties can be expressed by the moments of the probability density function and thus

---

[1]It is not clear, however, to what extent the independence assumption holds true. For example muscle fibers belonging to the same $\alpha$-motor neuron have a common drive, muscle spindle activity increases synchronicity between $\alpha$-motor neuron, etc. (see discussion in 1.2).

estimation is more robust and not prone to the curse-of-dimensionality [13]. This serves as motivation to study the probability density function of surface myoelectric signals.

To date, only a few studies have investigated this fundamental question and mostly assume the signals to follow a Gaussian density function [107, 72]. However, Clancy and Hogan showed that the signals density lies between a Gaussian and a Laplace probability density function [33]. In what follows we will briefly remind the functional forms of the two distributions.

The Gaussian probability density function for a single dimension is

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2.4}$$

where $\sigma$ is the standard deviation and $\mu$ the mean of the distribution. The multivariate Gaussian probability density for $d$ dimensions is

$$p(\vec{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp{-\frac{1}{2}\left(\vec{x} - \vec{\mu}\right)^T \Sigma^{-1} \left(\vec{x} - \vec{\mu}\right)} \tag{2.5}$$

where $\Sigma$ is the covariance matrix and $\vec{\mu}$ the $d$-dimensional mean. The functional form of the Laplace density is as follows. For the one dimensional case we have

$$p(x) = \frac{1}{2s} \exp{-\frac{|x-\theta|}{s}} \tag{2.6}$$

where $\theta \in (-\infty, \infty)$ and $s > 0$ are location and scale parameter, respectively. Estimators for these parameters are given by the first and second moment:

$$\theta = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{2.7}$$

which is the mean. The variance instead is [96]

$$2s^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \theta)^2. \tag{2.8}$$

The multivariate Laplace density for $d$ dimensions in turn is

$$p(\vec{x}|\Sigma) = \frac{2}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(\frac{\vec{x}^T\Sigma^{-1}\vec{x}}{2}\right)^{\frac{v}{2}} K_v\left(\sqrt{2\vec{x}^T\Sigma^{-1}\vec{x}}\right) \tag{2.9}$$

where $\nu = (2-d)/2$ and $K_\nu(\cdot)$ denotes the modified Bessel function of the third kind [96]. Further details about the Laplace distribution can be found in [96].

We compare the three models by computing the Kulback-Leibler divergence [36]

$$D_{\mathrm{KL}}\left(p_{\mathrm{m}}||p_{\mathrm{h}}\right) = E_{p_{\mathrm{m}}}\left(\log\left(\frac{p_{\mathrm{m}}}{p_{\mathrm{h}}}\right)\right). \tag{2.10}$$

The $D_{\mathrm{KL}}$ of two probability distributions is zero iff they are equal and positive otherwise. Because of this property the KL divergence is used to measure "closeness"[2] of probability distributions.

### 2.2.1.1 Is the distribution closer to a Gaussian or a Laplacian?

We find a clear trend towards Gaussianity for session 2 in table 2.2 when contraction strength is increased. This finding is in agreement with [114] where the authors compute the negentropy of MES which is a measure of non-Gaussianity and find it to decrease as contraction force is increased. However, we do not find the same trend for session 1. This deviation could come from the fact that no feedback of contraction force was given to the subject. It had to rely on its subjective perception of force level.

Among contractions there is no clear pattern of whether they are rather Gaussian or Laplacian. We conclude that surface electromyogram (sEMG) are neither clearly Gaussian nor Laplacian which is in agreement with [33].

| Session 1 | | | | |
|---|---|---|---|---|
| contraction | 20% MVC | 40% MVC | 60% MVC | all levels |
| hand open | 96 | 125 | 126 | 91.80% |
| hand close | 122 | 121 | 118 | 95.5% |
| flexion | 95 | 103 | 109 | 81.22% |
| extension | 113 | 103 | 101 | 83.86% |
| pronation | 116 | 94 | 103 | 82.8% |
| supination | 95 | 39 | 82 | 57.14% |
| abduction | 107 | 96 | 104 | 81.22% |
| adduction | 114 | 118 | 121 | 93.39% |
| all contractions | 85.12% | 79.27% | 85.71% | |
| **Session 2** | | | | |
| hand open | 50 | 83 | 112 | 64.81% |
| hand close | 32 | 53 | 101 | 49.21% |
| flexion | 35 | 93 | 101 | 60.58% |
| extension | 60 | 40 | 75 | 46.30% |
| pronation | 64 | 98 | 106 | 70.90% |
| supination | 98 | 104 | 114 | 83.60% |
| abduction | 94 | 91 | 87 | 71.96% |
| adduction | 75 | 108 | 116 | 79.10% |
| all contractions | 50.4% | 66.47% | 80.56% | |

Table 2.2: Gauss or Laplace? Both tables are based on recordings with the same experimental protocol and setup from the same subject but on different days.

---

[2]Note that the KL divergence is not symmetric in its arguments and, hence, is no metric.

### 2.2.1.2 Mixture model

Besides computing a Laplacian and a Gaussian probability density function we fit a mixture model of the two in order to test the hypothesis that the sEMG follows a distribution "between" Gauss and Laplace [33]. Means are set to zero $\mu = \theta = 0$ and thus we have

$$p(x|\sigma, s, \eta) = \eta p(x|\sigma) + (1 - \eta)p(x|s). \qquad (2.11)$$

with $\eta$ being the mixture weight. Fitting this probability density model is achieved by maximizing the likelihood via expectation maximization (EM) [13].

As to be expected, we find the mixture model to give the best match in terms of KL divergence. However, there are between 30 and 60 exceptions out of 1008 (number of electrodes times number of contractions) where the KL divergence is smaller for the Laplace model than for the mixture model. This is due to the EM algorithm stuck in a local minimum. However, we do not intend to correct this as those signals are few and all of them are weak and hence very likely irrelevant for any further investigation.

Even though the mixture model fits the distribution of some signals quite well there are still many signals that have an asymmetric probability distribution (see for example figure 2.2) which cannot be fit by it as both, Gauss and Laplace, distributions are symmetric. Asymmetricity can be quantified by the third moment, also termed skewness.

We observe that the tails of the skewed distributions follow a Gaussian on one side while on the respective other side a Laplacian. Unfortunately we do not know of a distribution model that captures this specific variability that we encounter in our measurements.

While the mixture model describes the peak of the MES it still cannot capture the skewness of the distribution. However, for those signals that have a symmetric distribution the mixture model presents a very good fit to the MES distribution (see for example probability density plots for contractions hand open, wrist flexion, pronation and supination in figure 2.1 and figure 2.2).

## 2.2.2 Skewness of myoelectric signals

To quantify the deviation from symmetricity seen in some signals by visual inspection of figure 2.2, we compute the skewness of the myoelectric signals. We then plot the skewness onto the electrode array grid as color plot in figure 2.3 and compare it to the standard deviation of the signal that quantifies the signal strength. This is done to investigate whether signal strength is correlated to skewness of the signal which for some contractions seems to be true as for example wrist extension and hand open but not for others as for example adduction. However, the overall correlation coefficient between skewness and standard deviation is 0.23, hence the correlation is weak.
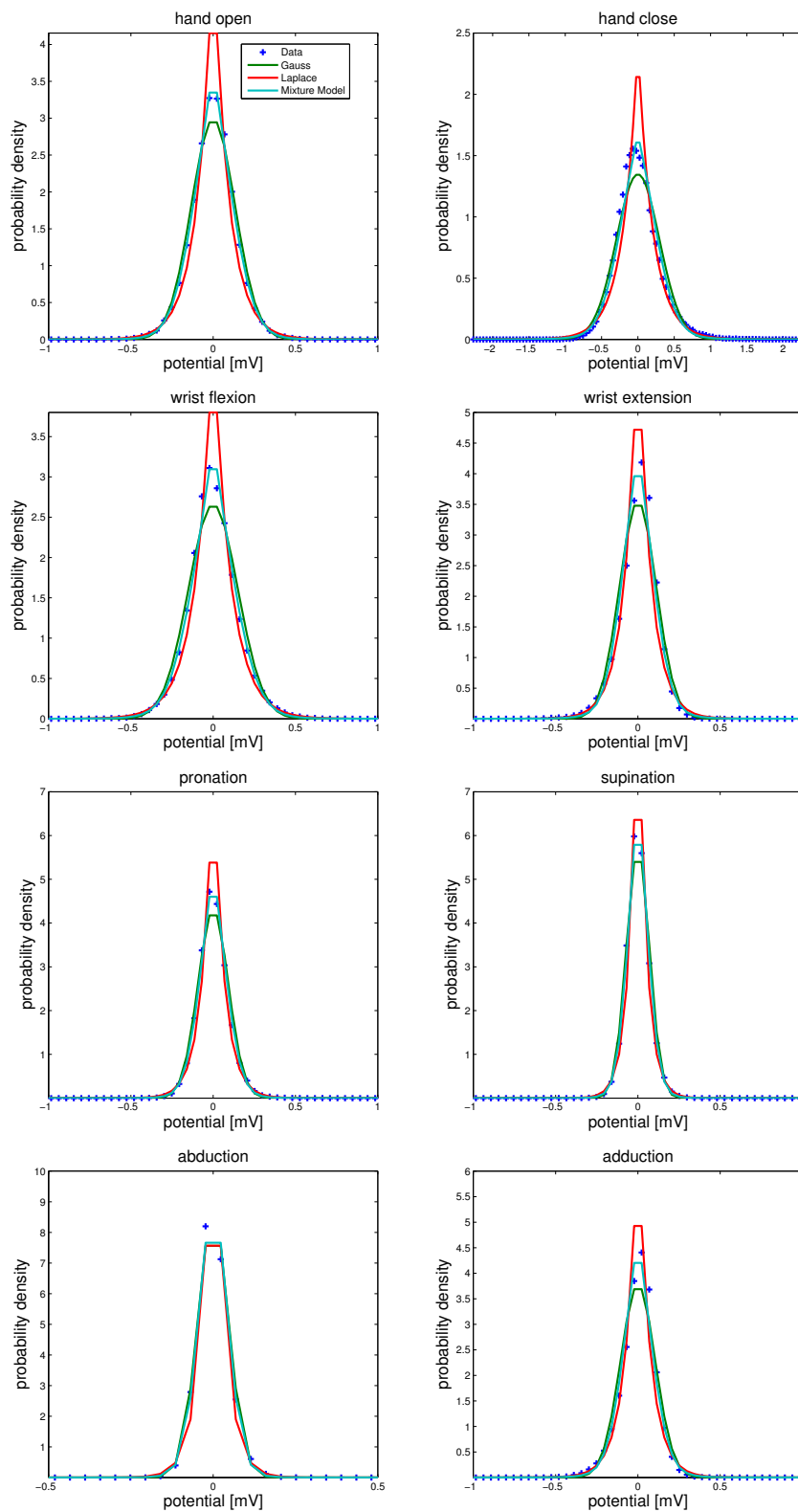
Figure 2.1: Probability density functions of all contractions of 60% MVC for channel 70.
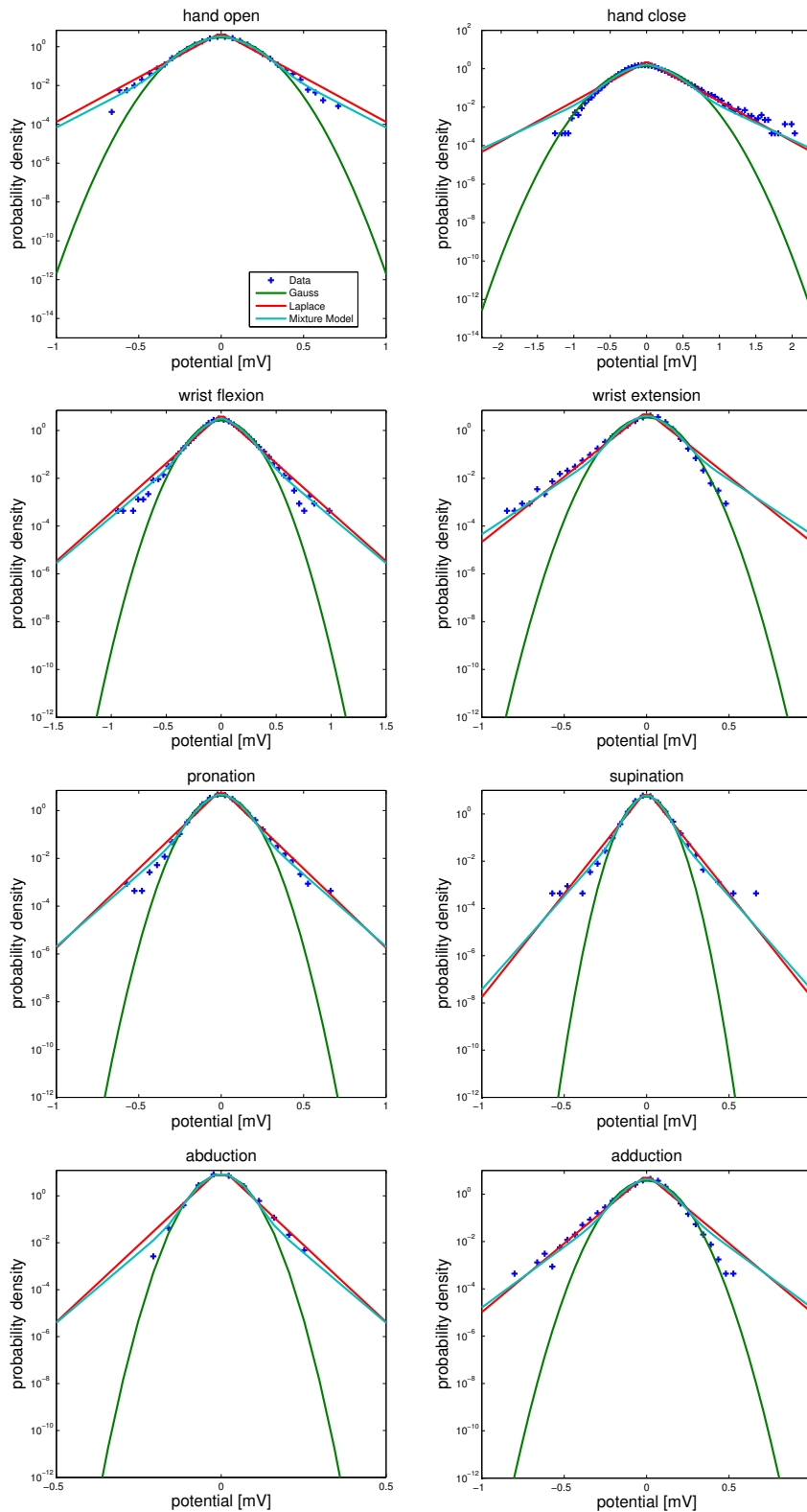
Figure 2.2: Semi-log plot of probability density functions of electrode 70. Blue crosses mark the empirical distribution, solid lines are probability density function fits: Gaussian (green line), Lapalce (red line) and mixture model (cyan line). Static contractions at 60% MVC.
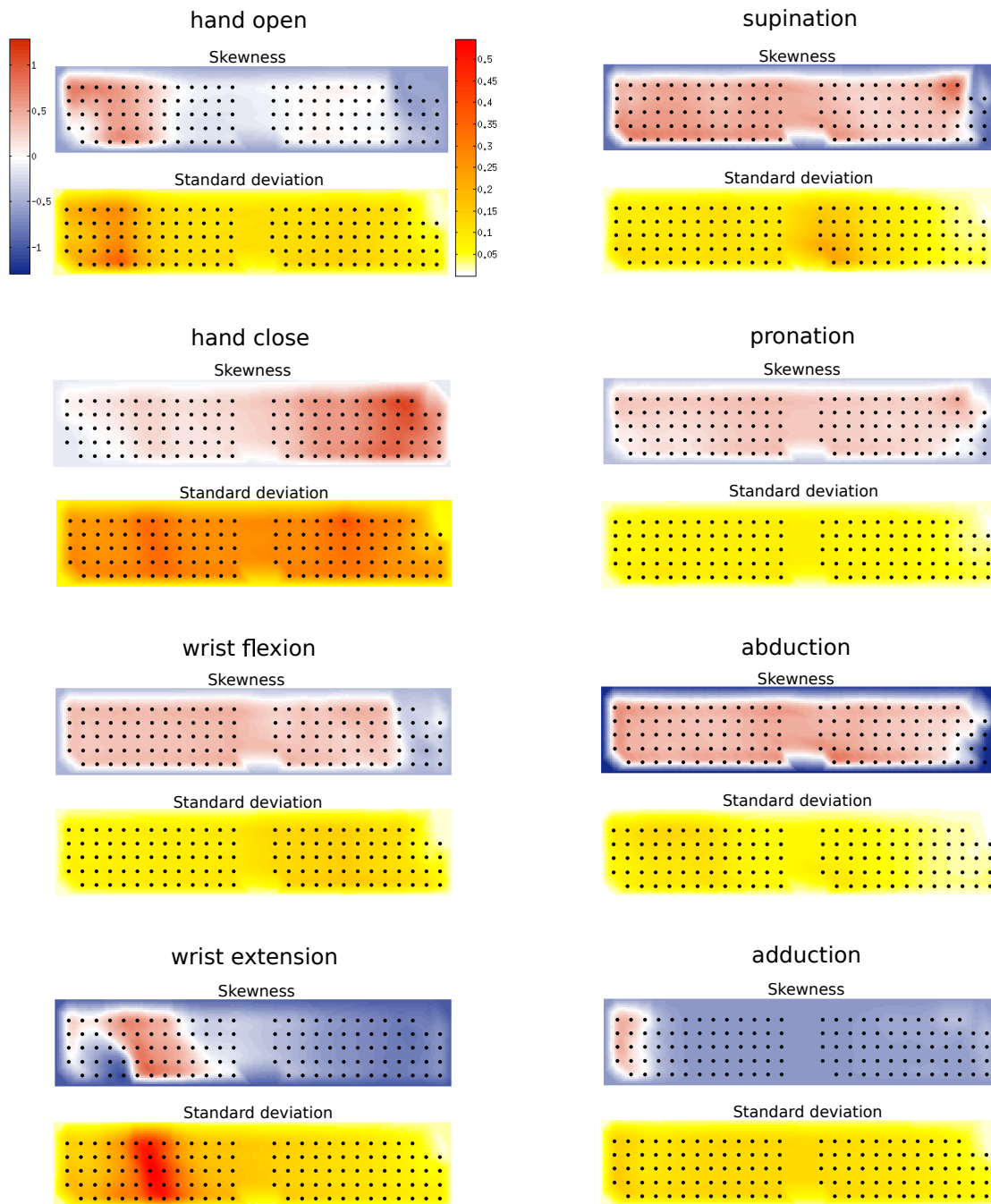
Figure 2.3: Skewness and standard deviation of MES recorded from two high density electrode arrays circumventing the lower limb. Contraction force at 60% MVC.

## 2.3 Bayesian filtering and state space models

Myoelectric signals are nonstationary whenever the contraction that is performed is dynamic or the force level is changing or the muscle is fatiguing. For most of the times an action is performed there will be nonstationary transients before the intended grip is achieved. The implication is that dealing with nonstationarity of the signal is unavoidable. However, as we see from table 2.1, only two studies dealt with MES in the nonstationary regime when investigating amplitude estimation. Indeed this situation holds true for most studies within the field of myoelectric prosthesis control based on pattern recognition systems as well.

However, now we present filtering methods designed for nonstationary inverse problems [85]. When inferring the force or the amplitude given the myoelectric signals, we have to deal with a genuine inverse problem. If the signal was stationary we could simply average over several measurements but given that force and amplitude may change in time we have to deal with a nonstationary inverse problem. A general approach to this kind of problems will be presented in the following which will give us a common framework to some of the previously presented studies, namely those employing Kalman filters [64, 49] as well as the Bayesian filter presented in [138].

We now describe the formal framework for Bayesian filtering for discrete time stochastic processes. Let $\{X_k\}_{k=0}^{\infty}$ and $\{Y_k\}_{k=1}^{\infty}$ be two stochastic processes. The random vector $X_k \in \mathbb{R}^{n_k}$ is called the state vector and represents the quantity that we are primarily interested in. The vector $Y_k \in \mathbb{R}^{m_k}$ represents the measurement. We refer to it as the observation at the $k$th time instant. The relation between stochastic events of the two processes shall obey the scheme

$$X_0 \to \quad X_1 \to X_2 \to \ldots \to \quad X_k \to \ldots$$
$$\downarrow \quad \downarrow \qquad\qquad\quad \downarrow$$
$$Y_1 \quad Y_2 \qquad\qquad\quad Y_k$$

which is an illustration of the following postulates for our stochastic processes:

1. The process $\{X_k\}_{k=0}^{\infty}$ is a Markov process, that is,

$$p(x_{k+1}|x_k, \ldots, x_0) = p(x_{k+1}|x_k). \tag{2.12}$$

2. The process $\{Y_k\}_{k=1}^{\infty}$ is a Markov process with respect to the history of $\{X_k\}$, that is,

$$p(y_k|x_k, \ldots, x_0) = p(y_k|x_k). \tag{2.13}$$

3. The process $\{X_k\}_{k=0}^{\infty}$ is connected to past observations $\{Y_k\}$ only through its own history, that is,

$$p(x_{k+1}|x_k, y_k, \ldots, y_1) = p(x_{k+1}|x_k).$$  (2.14)

If the stochastic processes $\{X_k\}_{k=0}^{\infty}$ and $\{Y_k\}_{k=1}^{\infty}$ satisfy conditions 1 to 3 then we call this pair an *evolution-observation model* [85]. To completely specify such a model we need the probability distribution of the initial state $X_0$, a discrete or continuous time model for the evolution of the (hidden) system state as well as a model for the measurement process. Indeed, the Kalman filter defines the simplest class of evolution-observation models, namely linear maps for hidden state evolution and measurement process with the stochasticity of evolution and measurement process is modeled by additive Gaussian noise [85]. However, in order to apply Bayesian filtering it is enough to model the respective probability distributions, namely the Markov transition kernels $p(x_{k+1}|x_k)$ (which can vary in time) and the likelihood $p(y_k|x_k)$.

In general we are interested in the estimation of the hidden state $x_k$ from previously measured data $\mathcal{D}_k = \{y_1, y_2, \ldots, y_k\}$, i.e. we look for $p(x_k|\mathcal{D}_k)$. Given the Markov assumption we find this probability distribution by a two-step procedure going sequentially through all data points from 1 to $k$:

**time evolution updating**:

$$p(x_{k+1}|\mathcal{D}_k) = \int p(x_{k+1}|x_k)p(x_k|\mathcal{D}_k)dx_k$$  (2.15)

**observation updating**:

$$p(x_{k+1}|\mathcal{D}_{k+1}) = \frac{p(y_{k+1}|x_{k+1})p(x_{k+1}|\mathcal{D}_k)}{p(y_{k+1}|\mathcal{D}_k)}$$  (2.16)

In order to evaluate the integral in equation (2.15) one often needs to resort to Monte Carlo sampling methods. The samples drawn to evaluate the marginalized probability distribution $p(x_{k+1}|\mathcal{D}_k)$ for the evolution state are called particles. This is where the term particle filter stems from and is a usual approach when no model of the hidden dynamics but only probability distributions are given, or the models are nonlinear and not linearizable. In case of a nonlinear model that is linearizable the linearized version is called an extended Kalman filter.

In the following we discuss another type of evolution-observation model as proposed in [138]. This gives a Bayesian filter where the dynamics of the hidden state are defined by a stochastic differential equation resulting in a nonlinear model that is not differentiable and hence not linearizable. However, Sanger proposed a Fokker Planck equation to model the time evolution updating step, hence no Monte Carlo sampling is needed to estimate $p(x_{k+1}|\mathcal{D}_k)$ (note also that this is therefore no particle filter).

## 2.3.1 Bayes filter with Fokker-Planck evolution step

The Bayesian filter model that we will study in detail in this thesis was proposed for amplitude estimation of myoelectric signals by T. Sanger in 2007 [138]. Here the hidden state dynamics have a time continuous description modeled by the stochastic differential equation

$$d\sigma = DdW + (U - \sigma)dN_\lambda. \tag{2.17}$$

The hidden state variable is $\sigma$, the standard deviation of the myoelectric signal (see also discussion in 2.1), $D$ is the "diffusion constant" of the Wiener process $dW$ (i.e. the differential of a Brownian motion), $U$ is a stochastic variable with uniform distribution between 0 and $\sigma_{\mathrm{MVC}}$, with $\sigma_{\mathrm{MVC}}$ being the maximum of the standard deviation as measured during a maximum value contraction. $U$ determines the jump size of the randomly occurring jumps modeled by a differential of a Poissonian counting process $dN_\lambda$ with rate $\lambda$. From this stochastic dynamics Sanger derived, under some approximations, the following Fokker-Planck equation

$$\frac{\partial p(\sigma, t)}{\partial t} = \frac{D^2}{2} \frac{\partial^2 p(\sigma, t)}{\partial \sigma^2} + \lambda(\frac{1}{\sigma_{\mathrm{MVC}}} - p(\sigma, t)) \tag{2.18}$$

which defines the evolution of the probability distribution of the hidden state, analogous to the Markov transition probability $p(x_{k+1}|x_k)$ in the Bayesian filter framework described above.

We introduced two minor changes into equation (2.18) with respect to [138], first we corrected[3] the factor of the first summand on the right hand side, the diffusion part, and second we introduced $\sigma_{\mathrm{MVC}}$ instead of demanding $\sigma$ to stay contained within $[0, 1]$. However, the first change does not affect the performance of the algorithm given that $D$ is simply a parameter that is to be optimized and has (to the current understanding) no specific physiological interpretation that could be tested quantitatively. Given the discrete nature of our data (measurements at a certain sampling rate) we discretize the Fokker-Planck equation and get

$$p_{t+\Delta t}(\sigma) = p_t(\sigma) + \Delta t \left( \frac{D^2}{2} \frac{\partial^2 p_t(\sigma)}{\partial \sigma^2} + \lambda \left( \frac{1}{\sigma_{\mathrm{MVC}}} - p_t(\sigma) \right) \right). \tag{2.19}$$

Finally the Bayes-Fokker-Planck filtering algorithm is

---

[3]Indeed, as $D$ is a parameter that will be optimized anyway the performance of the algorithm will not be affected by this mistake. This is true also for some other small mistakes in formulas in the original publication [138] which we corrected here.

Start with flat prior $p_0(\sigma) = \frac{\sigma_{\text{MVC}}}{m}$ and then do for each time step, i.e. each observation $s_{t>0}$

- **time evolution updating:** Discretizing the second derivative of equation (2.19) we obtain

$$p_{t+\Delta t}(\sigma) = p_t(\sigma) + \Delta t \left( \frac{D^2}{2} \frac{(p_t(\sigma + \Delta\sigma) - 2p_t(\sigma) + p_t(\sigma - \Delta\sigma))}{\Delta\sigma^2} + \lambda \left( \frac{1}{\sigma_{\text{MVC}}} - p_t(\sigma) \right) \right)$$
$$(2.20)$$

- **observation updating**

$$p_t(\sigma|\mathcal{D}_t) = \frac{p(s_t|\sigma)p_t(\sigma|\mathcal{D}_{t-\Delta t})}{p(s_t|\mathcal{D}_{t-\Delta t})}. \qquad (2.21)$$

- the maximum of the posterior $\sigma_t = \underset{\sigma}{\text{argmax}}\, p_t(\sigma|\mathcal{D}_t)$ gives the hidden state value, i.e. the myoelectric signal amplitude.

As model for the likelihood $p(s_t|\sigma)$ we will compare Gauss and Laplace distributions (see also 2.2). For the prior $p_t(\sigma|\mathcal{D}_{t-\Delta t})$ we use a histogram with $m$ bins set it to be flat at $t = 0$ as proposed by Sanger [138], i.e. $p_0(\sigma) = \frac{\sigma_{\text{MVC}}}{m}$ and the evidence $p(s_t|\mathcal{D}_{t-\Delta t})$ is not evaluated or modeled explicitly but at each update the posterior $p_t(\sigma|\mathcal{D}_t)$ is normalized with some constant $C_t$ such that $\frac{1}{C_t} \int p(s_t|\sigma)p_t(\sigma|\mathcal{D}_{t-\Delta t}) d\sigma = 1$. With this the algorithm is fully set and we may now look for optimal values for the parameters $D$ and $\lambda$. The additional parameter $m$ of the histogram prior fixes the resolution of the amplitude estimation within the interval $(0, \sigma_{\text{MVC}}]$. We decided to fix $m = 100$ which accounts for a high enough resolution for our purposes.

### 2.3.1.1 Conjugate prior

We will now extend the approach proposed in [138] by replacing the histogram prior with the so called conjugate prior distribution. Before we point out the advantages of doing so, we introduce the necessary basics.

Two distributions are conjugate if they belong to the same family of distributions. In Bayesian theory a prior is called conjugate if it belongs to the same family as the posterior. More specifically this means that the prior does not change its functional form when the Bayes formula is applied w.r.t. a specific likelihood distribution [102].

Practically, the conjugate prior for the variance of a Gaussian likelihood as well as of a Laplace likelihood with known mean (in our case $\mu = 0$ as stated earlier) is the same, namely the inverse gamma distribution [118]

$$p(\sigma^2|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left( \sigma^2 \right)^{-\alpha-1} \exp\left( -\beta/\sigma^2 \right) \qquad (2.22)$$

with a shape parameter $\alpha > 0$ and a scale parameter $\beta > 0$. $\Gamma(\cdot)$ is the gamma function [16]. For convenience we change notation for the variance $\sigma^2$, for the remainder of this chapter we use $v := \sigma^2$. Later we will need the first two moments of distribution (2.22), which are

$$E(v) \;=\; \frac{\beta\Gamma(\alpha-1)}{\Gamma(\alpha)} = \frac{\beta}{\alpha-1} \tag{2.23}$$

$$E(v^2) \;=\; \frac{\beta^2\Gamma(\alpha-2)}{\Gamma(\alpha)} = \frac{\beta^2}{(\alpha-1)(\alpha-2)} \tag{2.24}$$

where we have used the following property of the gamma function $\Gamma(x+1) = x\Gamma(x)$ [16]. Note that the gamma function $\Gamma(x)$ requires $x > 0$ hence the second moment requires $\alpha > 2$.

Note that using equation (2.22) as prior distribution will reduce the amount of parameters of the prior from $m-1$ to only two. Furthermore, due to the conjugacy we get simple closed form expressions of these two parameters for the observation update (2.21)

$$\check{\alpha}_t \;=\; \alpha_t + \frac{1}{2} \tag{2.25}$$

$$\check{\beta}_t \;=\; \beta_t + \frac{s_t^2}{2} \tag{2.26}$$

$\check{\alpha}_t$ and $\check{\beta}_t$ are the parameters of the posterior at time $t$ having observed the signal $s_t$.

Equation (2.22) has a support of $(0, \infty)$ while above we defined the support to be $(0, \sigma_{\mathrm{MVC}}]$, naturally limited by the finite amount of motor unit action potentials that can be fired at an instance in time. However, for the following derivation we set $\sigma_{\mathrm{MVC}} = \infty$. Given the exponential decay of the inverse Gamma distribution we do not expect this approximation to affect the filter performance significantly. Furthermore, the Fokker-Planck equation equation (2.18) was originally defined for the probability density of the standard deviation $\sigma$ rather than the variance $v$. Hence we either change the prior (2.22) to $p(\sigma|\alpha, \beta)$ or we redefine the Fokker-Planck equation in terms of the variance $v$. We decide to do the latter because this way we keep all equations as before and only need to reinterpret the result of the filter, i.e. we get the variance (from which we may compute the standard deviation $\sigma = \sqrt{v}$). Taking these two changes into account we rewrite equation (2.19) to

$$p(v|\alpha_{t+\Delta t}, \beta_{t+\Delta t}) = p(v|\alpha_t, \beta_t) + \Delta t \left( \frac{D^2}{2} \frac{\partial^2 p(v|\alpha_t, \beta_t)}{\partial v^2} - \lambda p(v|\alpha_t, \beta_t) \right). \tag{2.27}$$

Our aim is to find a closed form expression for the parameters also for the evolution updating. For this we need to find two equations of the parameters $\alpha_{t+1}$ and $\beta_{t+1}$ in dependence of these parameters at the previous time step $\alpha_t$ and $\beta_t$. We find these by computing the first and second moment of the propagated probability distribution

$p(v|\alpha_{t+\Delta t}, \beta_{t+\Delta t})$. For notational convenience we rewrite $p(v|\alpha_{t+\Delta t}, \beta_{t+\Delta t}) =: p_{t+\Delta t}(v)$ and $p(v|\alpha_t, \beta_t) =: p_t(v)$.

The first moment is given by

$$
\begin{aligned}
E_{p_{t+\Delta t}}(v) &= \int_0^\infty v p_{t+\Delta t}(v) dv \\
&= (1 - \lambda \Delta t) E_{p_t}(v) + \frac{D^2 \Delta t}{2} \left( \left[ v \frac{\partial p_t(v)}{\partial v} \right]_0^\infty - [p_t(v)]_0^\infty \right) \\
&= (1 - \lambda \Delta t) E_{p_t}(v) + \frac{D^2 \Delta t}{2} \left( \left[ \left( \frac{\beta_t}{v} - 1 - \alpha_t \right) p_t(v) \right]_0^\infty - [p_t(v)]_0^\infty \right) \\
&= (1 - \lambda \Delta t) E_{p_t}(v) \\
&= (1 - \lambda \Delta t) \frac{\beta_t}{\alpha_t - 1}
\end{aligned}
$$

(2.28)

where in the last step we used equation (2.23). The second moment is

$$
\begin{aligned}
E_{p_{t+\Delta t}}(v^2) &= \int_0^\infty v^2 p_{t+\Delta t}(v) dv \\
&= (1 - \lambda \Delta t) E_{p_t}(v^2) + \frac{D^2 \Delta t}{2} \left( \left[ v^2 \frac{\partial p_t(v)}{\partial v} \right]_0^\infty - 2 \int_0^\infty v \frac{\partial p_t(v)}{\partial v} dv \right) \\
&= (1 - \lambda \Delta t) E_{p_t}(v^2) + \frac{D^2 \Delta t}{2} \left( [(\beta_t - v - \alpha_t v) p_t(v)]_0^\infty - 2[v p_t(v)]_0^\infty + 2 \right) \\
&= (1 - \lambda \Delta t) E_{p_t}(v^2) + D^2 \Delta t \\
&= \frac{\beta_t^2 (1 - \lambda \Delta t)}{(\alpha_t - 1)(\alpha_t - 2)} + D^2 \Delta t
\end{aligned}
$$

(2.29)

where we used equation (2.23) and equation (2.24) in the last step. Assuming that the propagated distribution $p_{t+\Delta t}(v)$ is to a good approximation an inverse Gamma distribution, given a high sampling rate of 2048 and $\lambda$ and $D$ to be very small (see also 2.3.1.2), we can identify the parameters $\alpha$ and $\beta$ via the two moments as follows

$$
\frac{\beta_{t+\Delta t}}{\alpha_{t+\Delta t} - 1} = (1 - \lambda \Delta t) \frac{\beta_t}{\alpha_t - 1}
$$

(2.30)

$$
\beta_{t+\Delta t} = \frac{(1 - \lambda \Delta t) \beta_t}{\alpha_t - 1} (\alpha_{t+\Delta t} - 1)
$$

(2.31)

and

$$
\frac{\beta_{t+\Delta t}^2}{(\alpha_{t+\Delta t} - 2)(\alpha_{t+\Delta t} - 1)} = \frac{\beta_t^2 (1 - \lambda \Delta t)}{(\alpha_t - 1)(\alpha_t - 2)} + D^2 \Delta t
$$

$$
\beta_{t+\Delta t}^2 = \left( \frac{\beta_t^2 (1 - \lambda \Delta t)}{(\alpha_t - 1)(\alpha_t - 2)} + D^2 \Delta t \right) (\alpha_{t+\Delta t} - 2)(\alpha_{t+\Delta t} - 1)
$$

(2.32)

squaring equation (2.31) and equating with equation (2.32) we get

$$
\begin{aligned}
\frac{(1-\lambda\Delta t)^2\beta_t^2}{(\alpha_t-1)^2}(\alpha_{t+\Delta t}-1) &= \left(\frac{\beta_t^2(1-\lambda\Delta t)}{(\alpha_t-1)(\alpha_t-2)}+D^2\Delta t\right)(\alpha_{t+\Delta t}-2) \\
\frac{(\alpha_{t+\Delta t}-1)}{(\alpha_{t+\Delta t}-2)} &= \underbrace{\frac{\alpha_t-1}{(\alpha_t-2)(1-\lambda\Delta t)}+\frac{D^2\Delta t(\alpha_t-1)^2}{(1-\lambda\Delta t)^2\beta_t^2}}_{:=\zeta_t} \\
\alpha_{t+\Delta t} &= \frac{2\zeta_t-1}{\zeta_t-1} \qquad\qquad\qquad (2.33)
\end{aligned}
$$

and plugging this into equation (2.31) gives

$$
\begin{aligned}
\beta_{t+\Delta t} &= \frac{(1-\lambda\Delta t)\beta_t}{\alpha_t-1}\left(\frac{2\zeta_t-1}{\zeta_t-1}-1\right) & (2.34) \\
&= \frac{(1-\lambda\Delta t)\beta_t\zeta_t}{(\alpha_t-1)(\zeta_t-1)} & (2.35)
\end{aligned}
$$

Finally we get the following form for the evolution-observation algorithm

Start with prior parameters $\alpha_0 = 2.1$ ($\alpha > 2$) and $\beta_0 = 0.1$. For $t > 0$ the update steps are as follows

- **time evolution updating**

$$\alpha_{t+\Delta t} = \frac{2\zeta_t - 1}{\zeta_t - 1} \tag{2.36}$$

$$\beta_{t+\Delta t} = \frac{(1 - \lambda\Delta t)\beta_t\zeta_t}{(\alpha_t - 1)(\zeta_t - 1)} \tag{2.37}$$

with

$$\zeta = \frac{\alpha_t - 1}{(\alpha_t - 2)(1 - \lambda\Delta t)} + \frac{D^2\Delta t(\alpha_t - 1)^2}{(1 - \lambda\Delta t)^2\beta_t^2} \tag{2.38}$$

- **observation updating**

$$\check{\alpha}_t = \alpha_t + \frac{1}{2} \tag{2.39}$$

$$\check{\beta}_t = \beta_t + \frac{s_t^2}{2} \tag{2.40}$$

- the maximum or mode of the posterior

$$v_{\max} = \frac{\check{\beta}_t}{\check{\alpha}_t + 1} \tag{2.41}$$

gives the hidden state value estimate, the variance of the myoelectric signal. Taking the square root gives the amplitude estimate, i.e. $\sigma_{\max} = \sqrt{v_{\max}}$.

Due to time constraints a thorough quality assessment can only be given for the original version of the Bayes-Fokker-Planck filtering algorithm in this thesis. The same assessment for the extension of this algorithm by the introduction of a conjugate prior will be topic of future efforts. However, we wish to note that the derivation presented above is a first step in the direction of extending the Bayes-Fokker-Planck filtering algorithm to multiple dimensions. We plan to achieve this by introducing the inverse Wishart distribution which is the multi dimensional version of the inverse Gamma distribution [118].

### 2.3.1.2 Parameter optimization

In order to find the optimal parameters $D$ and $\lambda$ for the evolution update step (2.20) we compute the mean signal-to-noise ratio over all 126 electrode signals and all eight contractions as well as ten repetitions of each contraction. The contractions are performed at 60 % MVC and we take 4 seconds of each signal during which the contraction is held static and signal stationarity can be assumed (which is necessary to compute the SNR).

We find a regime of high SNR and set $\lambda = 10^{-30}$ and $D = 10^{-15}$ for the following evaluations. Note that even smaller $\lambda$ or $D$ do not improve the SNR substantially. Given the very small parameter values one could erroneously infer that they could be set to 0. While this would certainly lead to a very high SNR, it would be due to the filter not following the signal but, instead, after a brief initiation domination of the prior over the likelihood would occur and the maximum aposteriori will not change anymore resulting in a constant output signal.
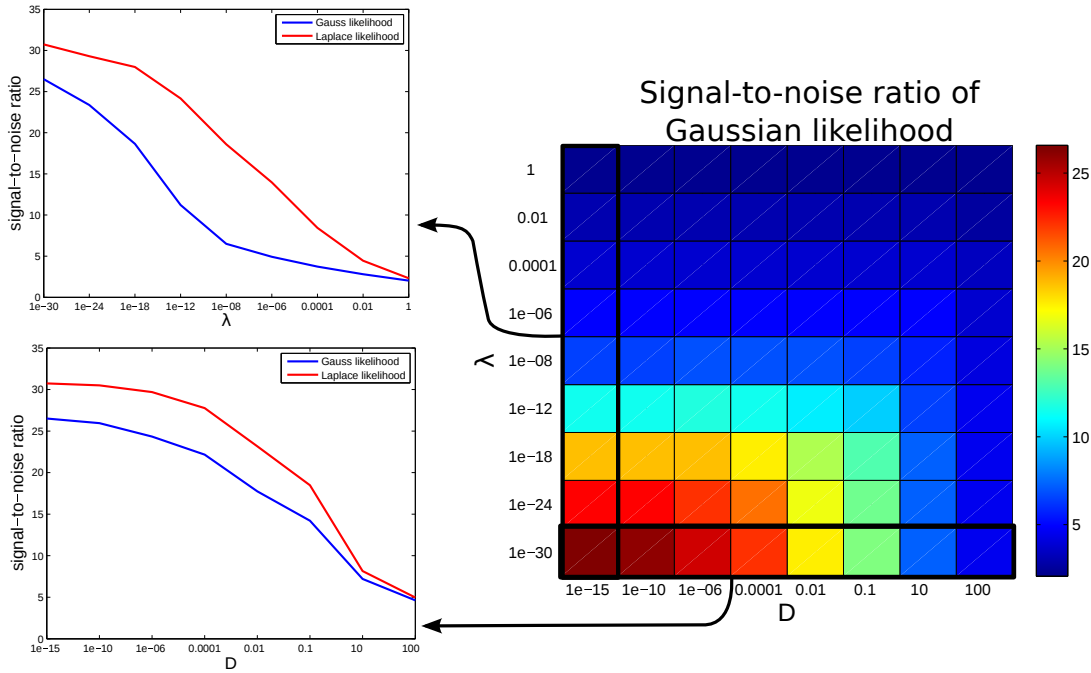


Figure 2.4: Color plot to the **right**: mean SNR over 10080 signals (126 electrodes, 8 classes and 10 repetitions) for the Bayes filter with a Gaussian likelihood in dependence on the two parameters $D$ and $\lambda$. Graphs on the **left**: comparison of mean SNR for two different likelihoods. Blue curves correspond to marked area in color plot, i.e. graphs for $D$ are produced with $\lambda = 10^{-30}$ and graphs for $\lambda$ are produced with $D = 10^{-15}$.

## 2.3.2 Results of Bayes-Fokker-Planck filtering

In the following we estimate the standard deviation of an artificially produced stochastic process with the Bayes-Fokker-Planck filtering method and compare it to a RMS estimation which, in this case, is the maximum likelihood estimator. We quantify the estimation quality by the RMSE. Afterward we compare Bayes-Fokker-Planck filtering to the state of the art filtering method [34, 52] using toolbox and data set provided

online by E.A. Clancy [27]. Finally we present filtering results of our own recordings of a dynamically varying contraction.

### 2.3.2.1 Artificial data

Before we filter real myoelectric signals we test the algorithm on a artificial signal. We produce the stochastic process which constitutes our artificial signal as follows. First a time series is produced by drawing from a standard normal distribution $\mathcal{N}(0,1)$ over 5000 seconds at a sampling frequency of 2048. This time series is then bandpass filtered with a Butterworth filter of order 7 and cut-off frequencies at 10Hz and 500Hz to restrict the frequency band to the one of a typical MES (the signal is normalized after filtering to have variance of 1 again). The amplitude, i.e. standard deviation, of the signal is drawn from a uniform distribution between 0 and 1.5 each time a jump occurs. Inter jump intervals are drawn from an exponential distribution with mean 500.

For the amplitude estimation quality, quantified by RMSE, we find 0.011 for the Bayes filter with a Gaussian likelihood while 0.041 for RMS with $T = 250$ms and 0.061 for $T = 100$ms. Hence the Bayes estimator is four to six fold better than RMS across reasonable (in terms of feasibility in prosthetic control application [54]) time window lengths.
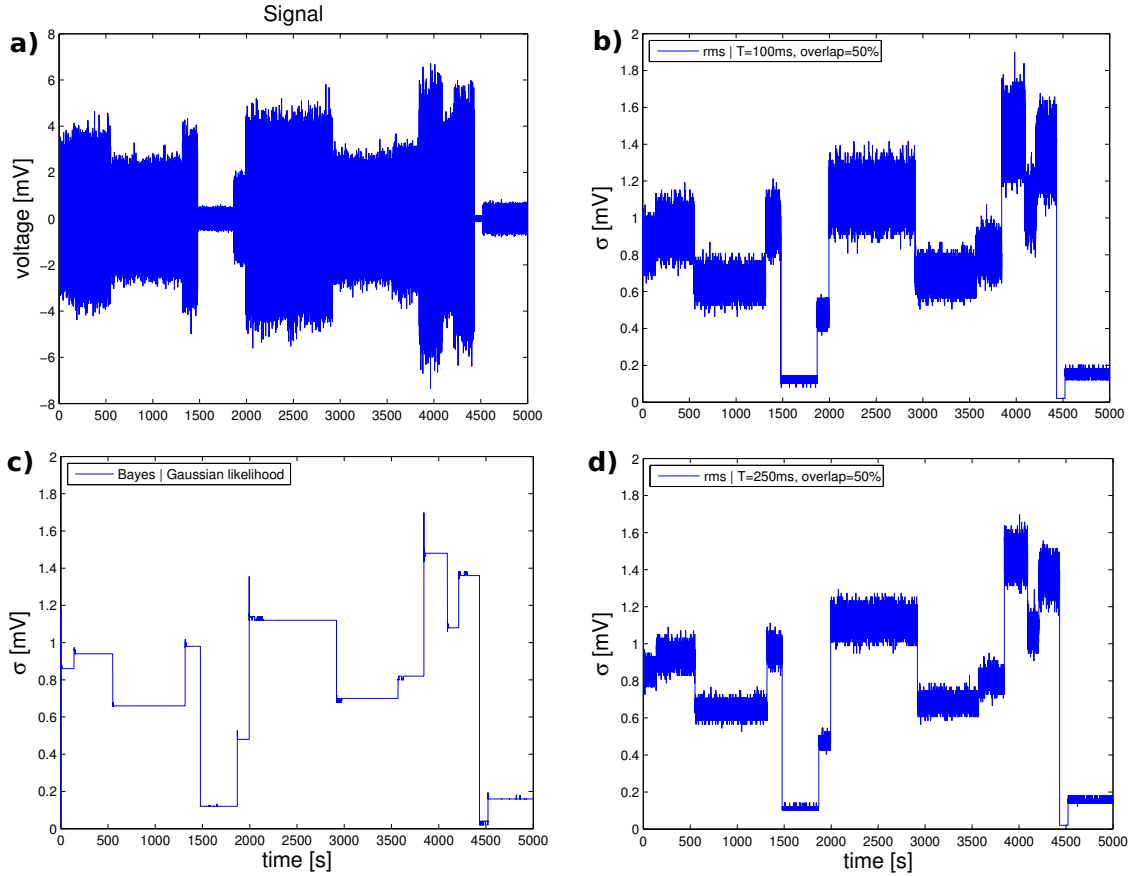
Figure 2.5: Performance comparison of RMS and Bayes filter on artificial MES. The artificial signal **a)** is a stochastic process resembling some characteristics of a MES, see text for a detailed description of how it is produced. **b)** and **d)** are the RMS of the artificial signal with 100ms and 250ms time window length, respectively, and 50% window overlap. **c)** is the Bayes filtered signal with a Gaussian likelihood model where $D = 10^{-16}$ and $\lambda = 10^{-30}$.

## 2.3.2.2 Comparison to state of the art amplitude estimation

In the following we compare the Bayes-Fokker-Planck filter with the state of the art filtering techniques [34] implemented in a toolbox provided by Clancy [27]. The myoelectric signal that is filtered is provided as part of the toolbox. We use the default parameter values of the toolbox assuming them to be optimized for this signal. We also compare the results to RMS amplitude estimation with different time window lengths and window overlaps.

As can be seen in figure 2.5 b) the amplitude estimation left and right of the peak is more flat as compared to c), d), e) and f). This comes from the whitening procedure that is applied by the amplitude estimation as proposed in [29]. Whitening of the signal

clearly also improves the performance of the Bayes-Fokker-Planck (BFP) algorithm as can be appreciated in figure 2.5 i) and j).

Moreover, we find that the Laplace likelihood gives a smoother estimate than a Gaussian likelihood, a feature that is certainly desirable for an amplitude estimation algorithm.
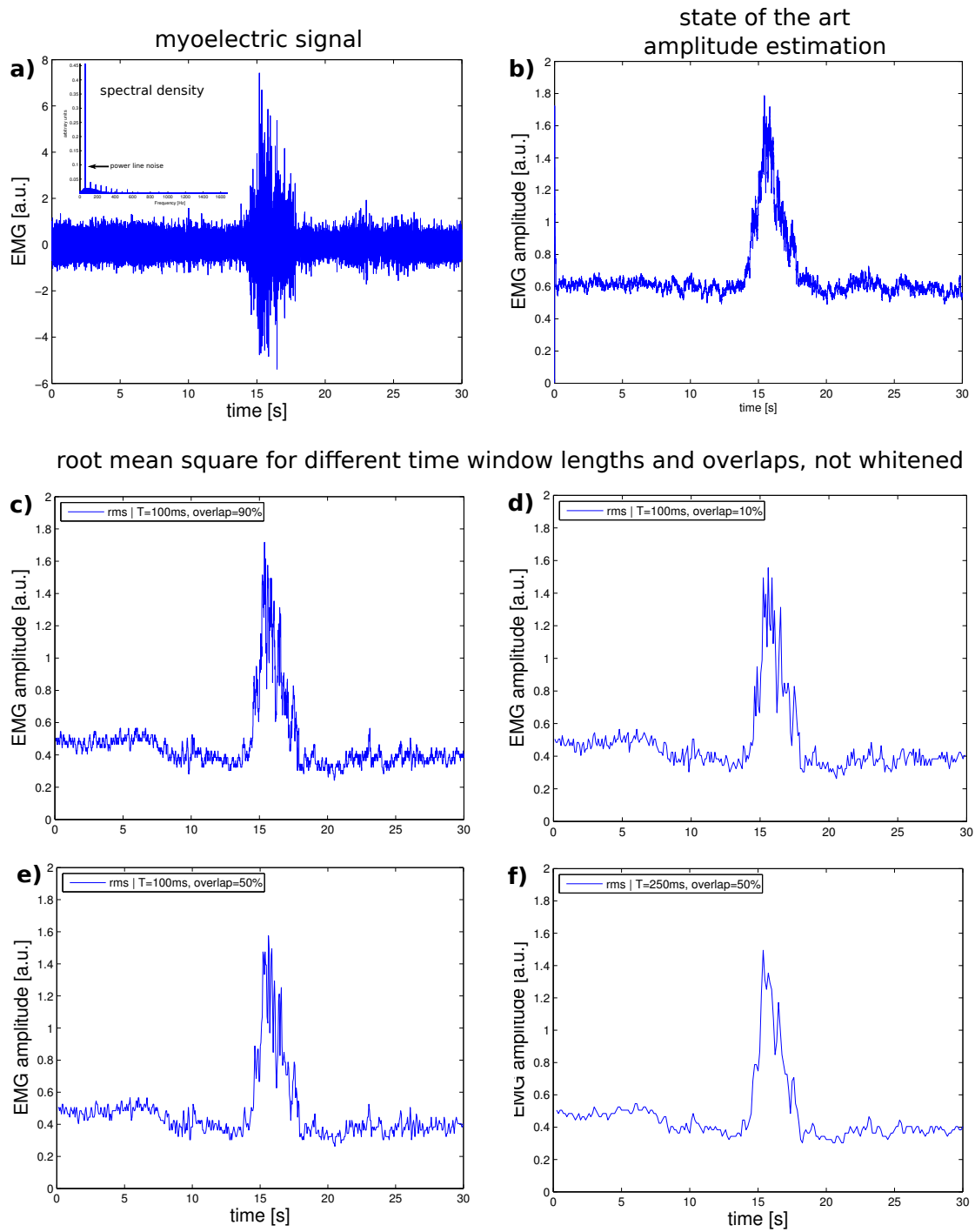
**a)** myoelectric signal

**b)** state of the art
amplitude estimation

root mean square for different time window lengths and overlaps, not whitened

**c)**

**d)**

**e)**

**f)**
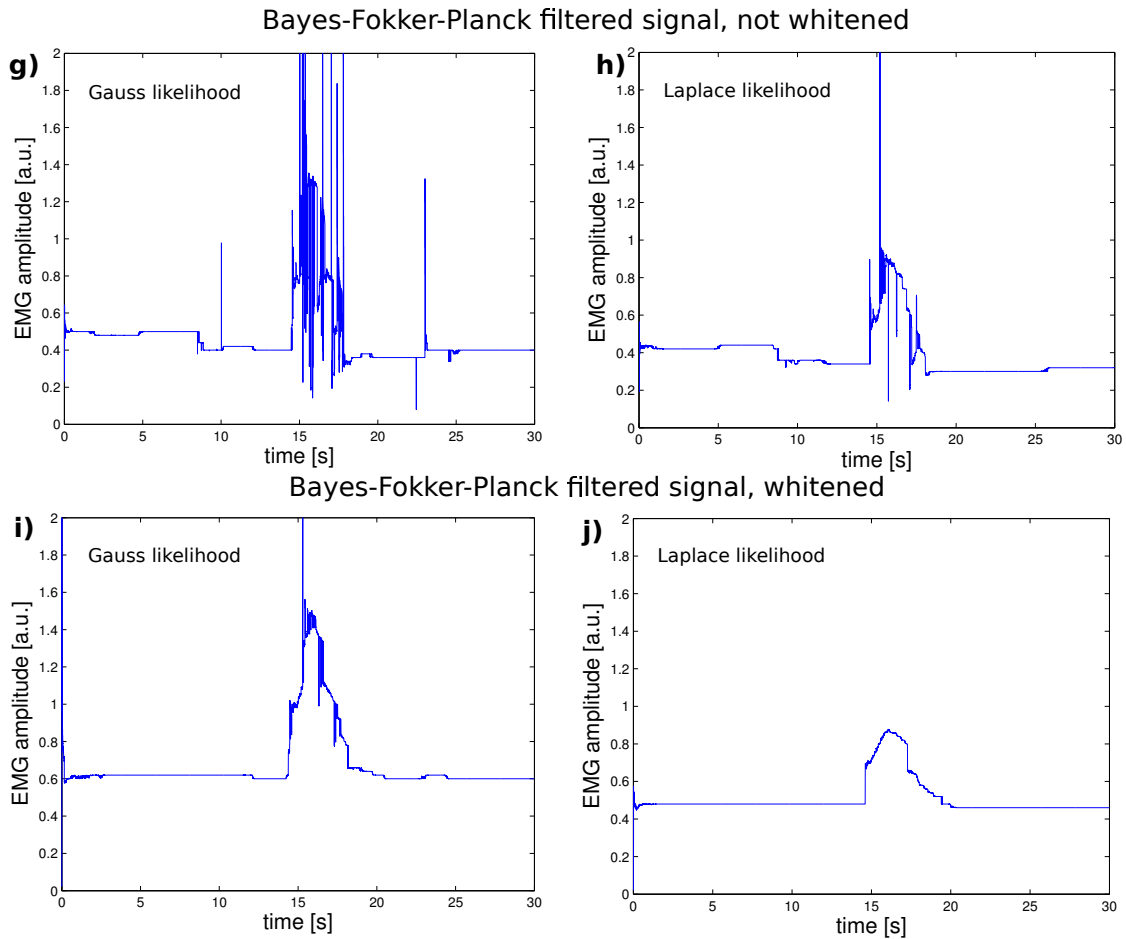
Figure 2.6

Figure 2.5: **a)** raw MES, inset is its spectral density that reveals a clear 60Hz (and its harmonics) power line noise. **b)** is the result produced by the toolbox provided by Clancy [27] using as "amplitude detector" the RMS, noise rejection and whitening all with default parameter values. **c)** to **f)** is the result of computing the RMS with different time window lengths and window overlaps. In **g)** and **h)** we find the result of filtering the raw signal with the Bayes-Fokker-Planck algorithm with two different likelihoods and $\sigma_{\mathrm{MVC}} = 2$, while in **i)** and **j)** we additionally whitened the same way it was whitened for **b)**.

### 2.3.2.3 Filtering at several contraction force levels

Here we test the performance of BFP on own data recordings. The myoelectric time series was measured while the subject was asked to perform contractions according to the following protocol: 4-8 seconds 20% MVC, 10-14 seconds 40% MVC, 16-20 seconds 60% MVC, 21-27 seconds ramp contraction from 0%-100% MVC, three short contractions of 20%, 40% and 60% MVC, respectively.

Furthermore, computing the SNR of all recordings where their mean amplitude is larger than 0.2 mV[4] during 60% MVC static contraction we find $7.4 \pm 1.7$ for BFP filtering with a Laplace likelihood and $6.8 \pm 2.0$ using a Gaussian likelihood. RMS with a time window of 200 ms gives in turn a mean SNR of $1.7 \pm 0.3$. Additionally, the best SNR achieved by RMS processing with a time window of 200 ms is 2.2 as compared to 2.8 which is the weakest SNR after BFP filtering with a Laplace likelihood.

We conclude that the BFP algorithm shows very good results on artificial data as well as off-line data recordings. Whether compared to standard root mean square amplitude estimation or more advanced estimation techniques, the BFP shows better performances in terms of signal-to-noise ratio. Also in terms of computational complexity the algorithm keeps up with other approaches tested, hence, BFP might be a very valuable choice for myoelectric applications where amplitude estimation is of importance.

---

[4]SNR for low amplitude signals BFP filtering gets extraordinary high ($> 25$). However, these signals are usually not relevant for application as they contain only a weak contribution of myoelectric signals.
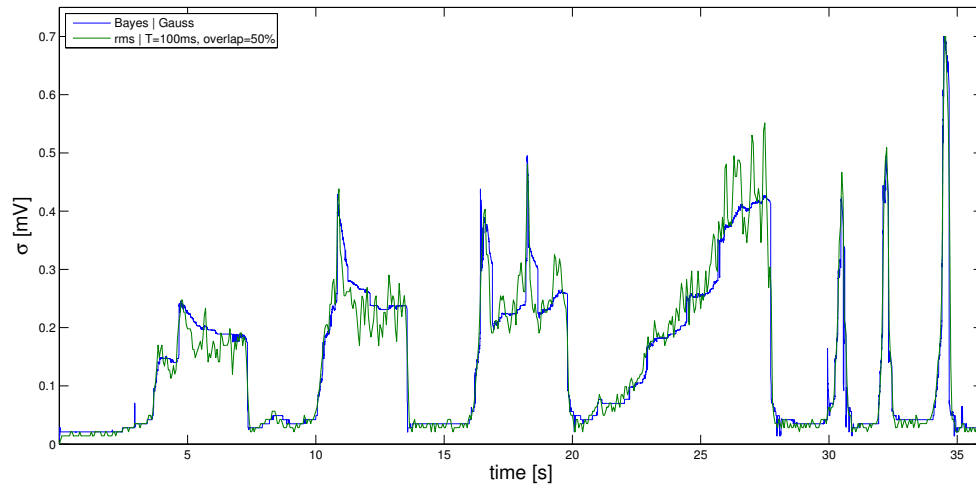
(a)
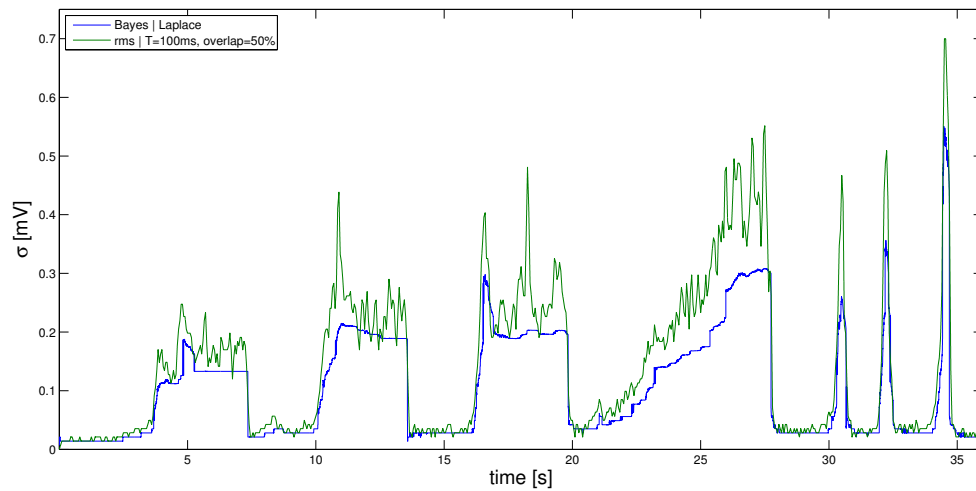


(b)

Figure 2.6: Comparison between RMS with a time window of $T = 100$ms and an overlap of 50% between windows and Bayes-Fokker-Planck filtering of the signal of electrode 1 during a hand close contraction. See text for explanation of the protocol. **a)** Gaussian likelihood function, **b)** Laplace likelihood function.

# 3 Machine learning approach

## 3.1 Basics

Generally, machine learning can be defined as follows [110]:

> We say that a machine *learns* with respect to a particular task $\mathbf{T}$, performance metric $\mathbf{P}$, and type of experience $\mathbf{E}$, if the machine reliably improves its $\mathbf{P}$ at $\mathbf{T}$ following $\mathbf{E}$.

The field of machine learning is usually [13, 22] divided into three main subfields of research, namely:

**Supervised learning** where the task $T$ is to infer a target or dependent variable $y$ given a vector of features or independent variables $\vec{x}$[1]. If $y$ is discrete, $T$ is called a classification task, if $y$ is continuous it is called a regression task (note that we stick to the usual notation where $y$ is one-dimensional, however, $y$ may be a vector). The performance metric $P$ is usually called loss, risk or error function and quantifies the deviation of the estimated $\hat{y}$ from the true $y$. The experience $E$ is normally a set of pairs $\{y_i, \vec{x}_i\}$. However, more recently approaches drew interest of the machine learning community where unlabeled data, i.e. data where no target value $y$ is given, is used to improve the performance of a genuinely supervised learning task. These kind of approaches are therefore called semi-supervised learning.

**Unsupervised learning** where $T$ is to extract statistical structure from (unlabeled) data. Examples are the task of finding clusters in the data as for instance through K-means clustering [105, 13] and mixture models [60] or we wish to find the independent sources signals as in independent component analysis (ICA) [79]. Depending on the specific task, $P$ may be the likelihood of the data $p(\{\vec{x}_i\}|\vec{\theta})$ telling how likely it is that the observed data is drawn from an assumed model distribution at a specific point $\vec{\theta}$ in parameter space. Another example for $P$ is the entropy $H(X)$ quantifying independence as in the maximum information ICA algorithm [6, 7].

---

[1]Note that the terminology target/features is mainly used in the machine learning community, whereas in the statistical learning community more often the terminology dependent/independent variables is used. Additional confusion may be produced by the slightly different use of the term features in the myoelectric pattern recognition community: there a feature denotes usually a function of the raw MES as for example the RMS, hence different features are different functions rather than the different input dimensions to the pattern recognition system.

**Reinforcement learning** where the general setting is quite different. Here the learning machine is an agent that can act in an environment and gets rewards or punishments from the environment. The task $T$ is to shape its perception-action policy such that it maximizes the obtained reward and/or minimize punishment, both given by numerical values. The performance metric $P$ is the reward or punishment for a certain action given a specific state of environment and agent. $E$ are pairs of action and reward [150]. However, we shall not be concerned with this category of machine learning algorithms in what follows.

Supervised and unsupervised learning can be seen as subfields of the mathematical field of statistical learning [60, 154]. In this thesis we will deal with the application of supervised and unsupervised learning techniques. While the problem at hand is genuinely a supervised learning problem, we will also employ unsupervised learning techniques for preprocessing the data in order to improve the performance achieved by the supervised learning algorithm.

A supervised learning algorithm finds a function, often called hypothesis,

$$\hat{y} = h_\theta(\vec{x})$$

out of a family of functions parameterized by $\theta$, which maps $\vec{x}$ onto $\hat{y}$ where ideally $\hat{y} = y$. The (locally) optimal parameters $\theta^*$ is optimal w.r.t. the performance measure $P$ given some data $\mathcal{D} = \{y_i, \vec{x}_i\}$. This optimization procedure is called training. However, if a function family has a high VC dimension [154], which is to say a high complexity, overtraining also called overfitting is likely to happen. Overfitting means a considerable drop in performance for out-of-set samples (i.e. previously unseen data). One speaks also of poor generalization ability of the hypothesis $h_{\theta^*}$ in such a situation [154]. A simple method called cross-validation allows to detect overfitting and thus compare the generalization ability of various models (families of functions). The idea is illustrated in *figure 3.1*.

## 3.2 Scheme of pattern recognition approach

### 3.2.1 Preprocessing

This step may consist of filtering in the spatial or temporal domain or both. In this work we dedicate an extensive study about spatial filters. Given the high density electrode array measurements we have a convenient setting for thoroughly studying the effect of spatial filters on classification accuracy. Chapter 3.3 will be dedicated to this topic.

Additionally temporal filtering may improve classification accuracy. We have studied in 2.3 a temporal filtering method for non-stationary myoelectric signals. This may be applied for problems related to amplitude estimation, such as force or torque estimation but it can equally well be used as preprocessing for pattern recognition approaches.
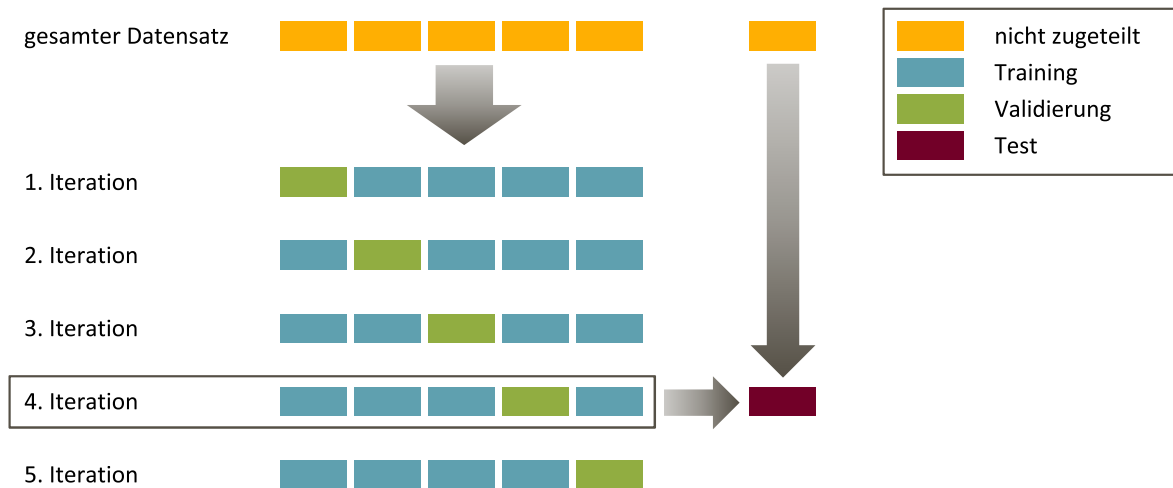
Figure 3.1: In its simples setting, model selection via cross-validation generally involves two data sets: a training set that is used for parameter estimation, and a validation set that is used for model selection [148]. The validation set provides a means of comparing the generalization ability of various models and detecting overfitting: if the performance on the validation set is significantly poorer than the performance on the training set, the model exhibits overfitting. If many models are compared with the validation set, the performance of the best selected model on the validation set is an optimistic estimate of the prediction performance on new data; in a sense, learning is then performed on the validation set. A third set called test set, distinct from the training and validation set, is then needed to assess the final performance of the model (adapted from [137]).

Please note that in this case terminology might get fuzzy. Sometimes temporal or spatial filtering might be called feature extraction. For example when we estimate the standard deviation, which is a common feature for myoelectric signal used pattern recognition approaches, by the BFP filter. The same ambiguity is also found for ICA algorithms such as information maximization [6] which might be seen as producing spatial filters for preprocessing or as feature extraction algorithm [7]. However one prefers to categorize these procedures will not limit the generality of our results or change our understanding, therefore we leave it as a matter of taste to the reader.

## 3.2.2 Features

In the field of machine learning a feature simply means an input dimension to a supervised learning algorithm. For unsupervised learning algorithm the term feature is sometimes used in relation to the statistical structure the algorithm extracts from the data but a precise definition is not established in the field of machine learning.

In the field of myoelectric pattern recognition the term feature mostly refers to a function of the measured signal as opposed to the raw signals themselves. A popular set of features used in myoelectric pattern recognition approaches is the mean absolute value, zero crossing, slope sign change and waveform length, commonly known as the Hudgins feature set [77].[2] Instead of the mean absolute value the root mean square is often calculated. Indeed, the latter is a very common feature for pattern recognition of myoelectric signals and also, as we have seen earlier, used to estimate the EMG amplitude. However, many features were tested for classification improvement. An early comparative study finds EMG histogram to be the best feature for a pattern classification approach [159]. In [14] 19 features were compared and evaluated about noise tolerance and computational complexity besides classification performance. The authors found that wavelet coefficients and cepstrum coefficients performed best. Apart from the features compared in these studies, investigations were also done on using higher order statistics as skewness and kurtosis [114] for classification. The authors showed that kurtosis together with variance lead to good results while skewness leads to poor performances.

Despite the many features tested, features relating to the EMG energy as mean absolute value and root mean square are still the most widely spread. In our pattern recognition system we use only the root mean square to keep our approach simple and the results comparable.

Normally, the calculation of a feature takes a set of data samples defined by a time window. When computing the root mean square longer time windows lead to improvements in classification accuracy but at the same time increase the controller delay of a hand prosthesis. A standard time window for RMS calculation is around 200 ms [45, 67].

---

[2]In the original paper Hudgins, et al. [77] used also the slope of the mean absolute value, but this is, however, often not considered part of the classical feature set [45, 67].

Investigations on the maximal controller delay have shown that it should not exceed 300 ms [54, 45]. However, in this thesis we investigate the effect of time window length for feature calculation on classification performance extensively.

### 3.2.3 Feature selection

Feature selection can be a means to deal with situations where the amount of features, i.e. dimensions, outnumbers the amount of data points [117, 66]. In some cases feature selection is superior to regularization to avoid overfitting (even though regularization will result in close to 0 coefficients for irrelevant features) [117]. Additionally feature selection can be applied to reduce the computational load of processing a large amount of features.

In this study, however, we do not have to deal with these issues, instead we use feature selection to address the question about electrode positioning and the minimum number of electrodes needed to achieve a high enough performance. A detailed discussion and results are presented in 3.4.

To prevent confusion about terminology we wish to emphasize that feature selection refers to the selection of input dimensions in accordance with the machine learning definition of features. It is not concerned, however, with the selection of good features according to the terminology within the myoelectric signal community where a feature refers to a function of the raw signals. Indeed the only feature we use throughout our investigations is the RMS.

### 3.2.4 Classification

Many classifiers have been investigated and compared within the myoelectric signal community. Multi layer perceptron [77, 25], linear discriminant analysis [45, 24, 67], support vector machines [119, 104], Gaussian mixture model [76] and hidden Markov models [21] to name a few.

However, the most popular due to its simplicity and good performance is linear discriminant analysis (LDA) which is the Bayes optimal classifier given the signals are distributed according to a Gaussian with equal covariance matrix for all classes.

### 3.2.5 General set-up for classification in this thesis

For the sake of convenience we wish to define here the standard classification scheme that is used if not explicitly stated otherwise.

Feature is *root mean square*, channel selection is *sequential forward selection* and the classifier is *linear discriminant analysis*. All these methods are simple standard methods used in the field and provide thus a good base to compare our results to other studies.

Furthermore, the data that we use (if not stated otherwise) is measured during static contractions of 60% maximum value contraction (MVC). And we always perform a 10-fold cross-validation to prevent overfitting.

### 3.2.6 Experimental setup

The results of this thesis are based on measurements taken with two high density electrode arrays wrapped around the upper forearm as depicted in figure 3.2. Each array consists of 64 Ag/AgCl electrodes with 8 mm inter-electrode distance. Conductive gel is applied to improve signal quality. We removed the power-line interference with a 45-55 Hz band stop (2nd order Butterworth), additionally the signals were filtered with a high pass whose cutoff is at 500 Hz (4th order Butterworth) and a low pass with cut-off frequency at 20 Hz (4th order Butterworth).

We investigate three able bodied subjects who perform repeatedly the following contraction procedure: 0.5 second contraction onset followed by 3 seconds of static contraction and then 0.5 second contraction offset. This pattern is repeated for 20%, 40% and 60% MVC with 2 seconds of rest in between. The same temporal procedure is performed for the eight different static contractions listed and illustrated in figure 1.1, namely hand open, hand close, wrist flexion, extension, supination, pronation, abduction, and adduction.
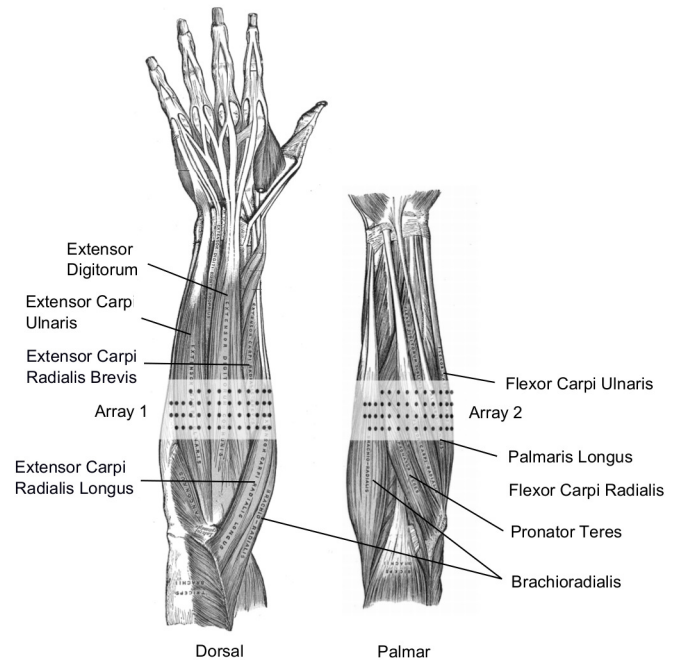


Figure 3.2: High density electrode array positions on the upper forearm [67].

## 3.3 Spatial filtering

Spatial high-pass filter have been successfully applied to surface myoelectric signals for the investigation of single motor unit activities [43, 50, 53, 121, 122]. Spatial filtering also improves MES force estimation as shown by Staudenmann et al. [145]. They showed that filtering based on principal component analysis (PCA) outperforms all data-independent filters tested, among those the Laplace filter. Furthermore, they find that using monopolar signals (i.e. unfiltered) results in better amplitude estimations than after applying a bipolar filter. More recently it was shown that spatial filtering also improves classification accuracy of upper limb contractions. Huang, et al.[75] investigated the classification of 15 different upper limp contractions following targeted muscle reinnervation with 128 monopolar electrodes. They used 11 different data-independent

spatial filters and found that double differential filter performed best if 12 channels are used for classification. Channels were selected by sequential forward selection (see 3.4 for explanation of the algorithm). Another study conducted by Hahne, et al.[67] recorded 96 monopolar MES using high density electrode arrays and filtered the signal using the data-dependent, supervised filter common spatial patterns (CSP). The authors showed that the resulting signals were superior in terms of discriminability by LDA to signals resulting from a longitudinal bipolar spatial filter and the computation of the Hudgins feature set.

They have shown that principal component analysis improves force estimation [145] and outperforms a Laplace spatial filter as well as what they call "optimally aligned multiple bipolar directions" which is the high density electrode array chunked into diagonal bipolar channels. Interestingly they also find that conventional bipolar configuration is worse then monopolar.

ICA has found application in EMG analysis for the extraction of action potentials of muscular units from surface myoelectric signals [113, 61], the detection of muscle fatigue [149], and the identification of the activity of individual muscles [112]. Furthermore, it was applied to improve classification accuracy of a pattern recognition system for myoelectric signal classification [111].
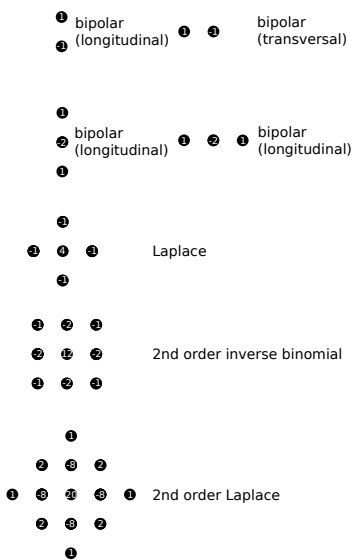
## 3.3.1 Data-independent filter

Data-independent filters are a set of predefined filter masks as shown in figure 3.3. We use the one proposed in [75] to be able to compare results.

Each filter mask is convolved with the electrode array where boundary conditions are set such that longitudinally we pad with zeros while transversally channels that are confining from the other electrode array are taken, hence we assume no boundary in this direction which is reasonable given that the entire arm is wrapped by two electrode arrays.



Figure 3.3: Filter masks for data-independent filters.

### 3.3.1.1 Results of spatial filtering with data-independent filters

In comparison the bipolar longitudinal filter mask performed best. Hence, we will compare data-dependent filters only with this data-independent filter mask.

Figure 3.4: Comparison of bipolar filter masks with unfiltered classification accuracy. **Upper** graphs: Accuracy over time window length, classification is done with the two best channels. **Lower** graphs: Accuracy over number of channels, time window is set to 20 ms. **Left** is static regime only while **right** figures result from including the dynamic regime.
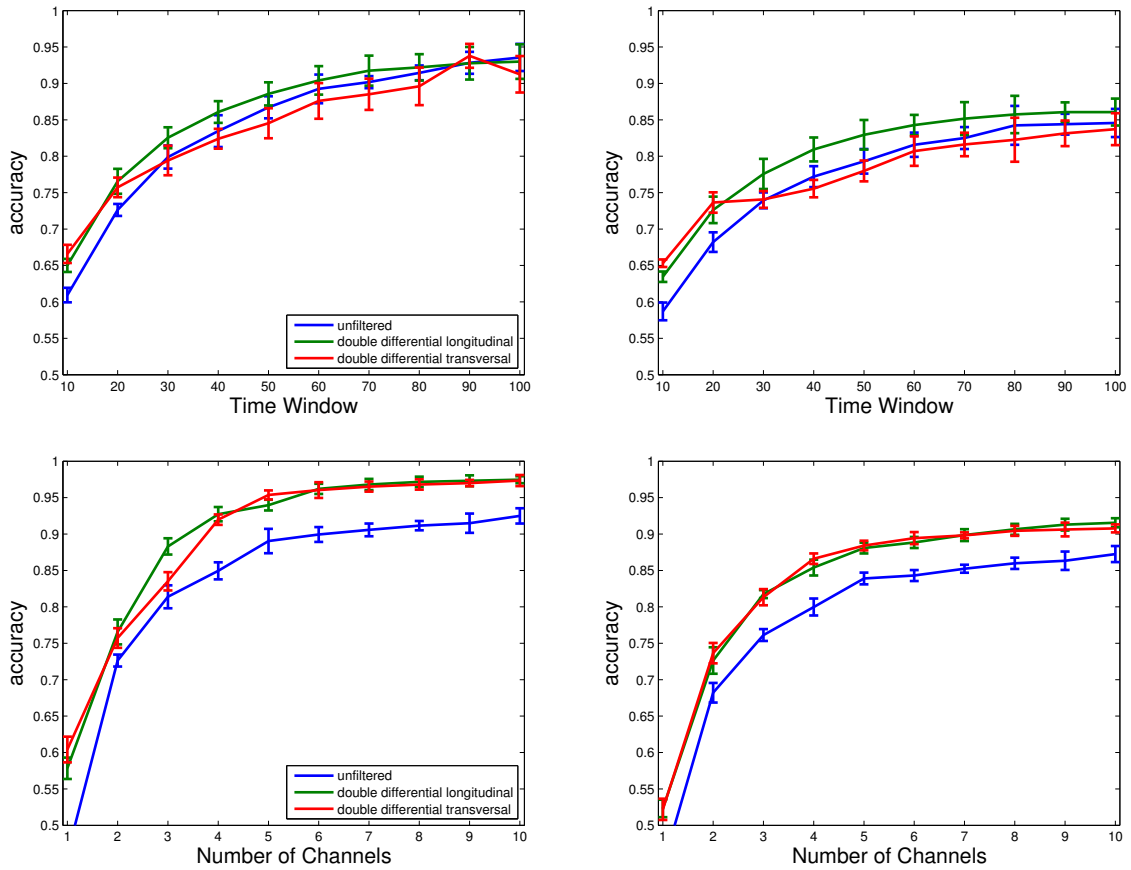
Figure 3.5: Comparison of double differential filter masks with unfiltered classification accuracy. **Upper** graphs: Accuracy over time window length, classification is done with the two best channels. **Lower** graphs: Accuracy over number of channels, time window is set to 20 ms. **Left** is static regime only while **right** figures result from including the dynamic regime.

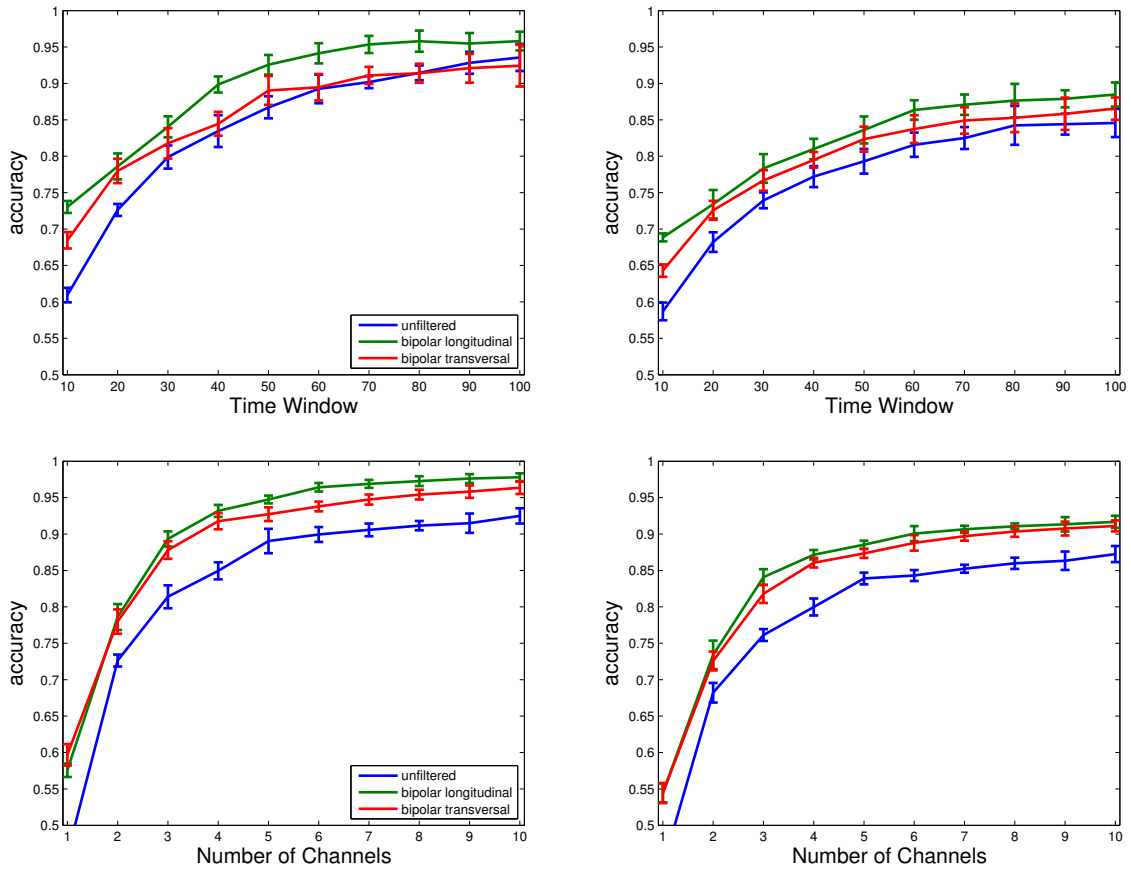Figure 3.6: Comparison of several data-independent filter masks with unfiltered classification accuracy. **Upper** graphs: Accuracy over time window length, classification is done with the two best channels. **Lower** graphs: Accuracy over number of channels, time window is set to 20 ms. **Left** is static regime only while **right** figures result from including the dynamic regime.
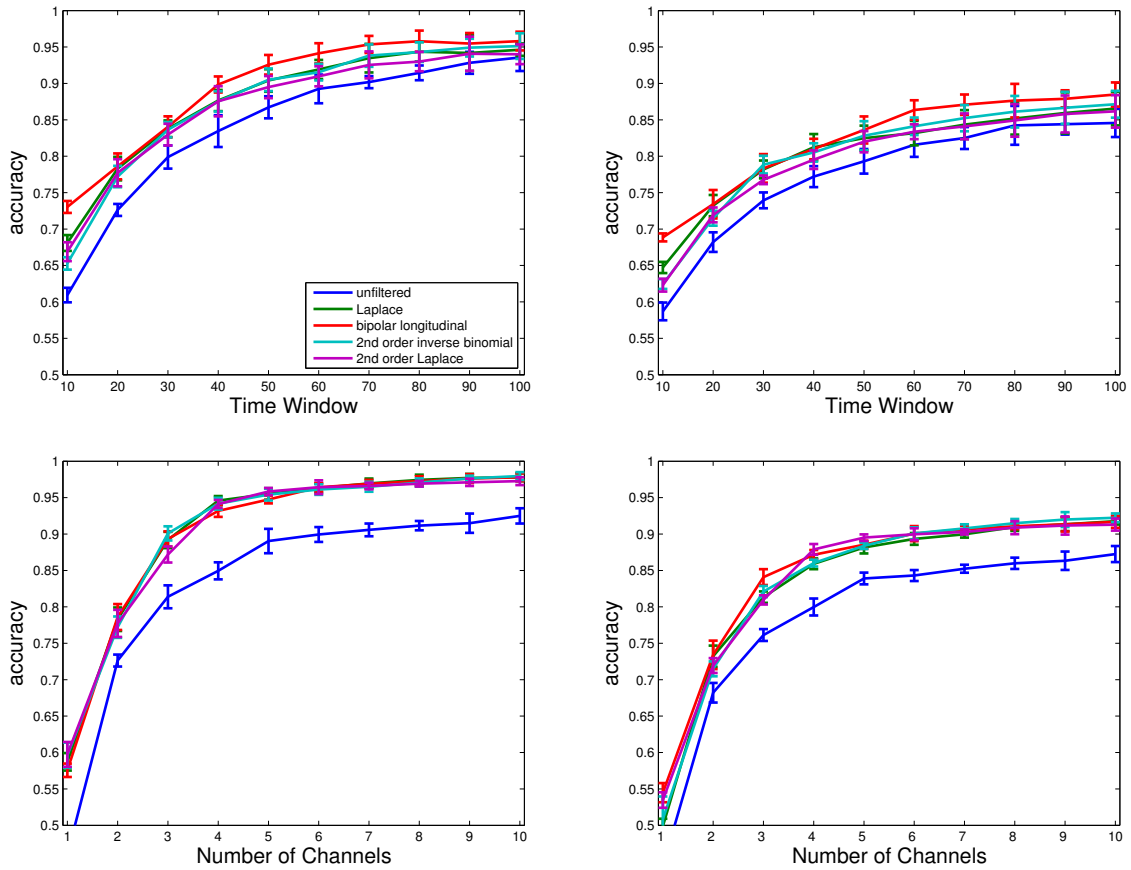
## 3.3.2 Data-dependent filter

In the sequel we present briefly the different algorithms that were used for data driven spatial filtering.

### 3.3.2.1 Principal component analysis

Principal component analysis a classical technique in statistical data analysis and is counted to the family of unsupervised learning algorithms [60]. It is commonly used for dimensionality reduction, lossy data compression, feature extraction and data visualization [84].

PCA was first introduced by Karl Pearson in 1901 [126]. By drawing a conceptual analogy to principal axes in mechanics, he aimed at extending the regression analysis where only the dependent variable is assumed to be stochastic while the dependent variables are not to the situation where both are stochastic. He approached the problem by minimizing the mean squared distance between data points and their projection onto the principal axes. Later in 1933, PCA was reinvented by Hotelling [74] which is why it is sometimes also called the Hotelling transform. He used a different approach leading to the same algorithm; he defined an orthogonal projection of the data onto a lower dimensional linear space, known as the principal subspace, such that the variance of the projected data is maximized. Hotelling himself named it the methods of principal components [74]. Furthermore, PCA is sometimes also called Karhunen-Loève transform which is the empirical version of the Karhunen-Loève theorem [84] which is a representation of a stochastic process as an infinite linear combination of orthogonal functions.

In short, PCA an orthogonal linear transformation where the first basis vector (eigenvector of the covariance matrix with largest eigenvalue) points in the direction of highest variance and all the following are orthogonal to the former basis vectors and point in the direction of . Given that the resulting principal components (eigenvectors) are orthogonal and given that the method is based on the covariance, the projected signals are decorrelated, whitened and in case of probability distribution that are completely characterized by the first and second moment, as for instance the Gauss and symmetric Laplace distributions, removing pairwise correlations makes signals independent. somewhere, probably in Bishop or ICA stated what the assumptions for PCA are.

In the field of myoelectric signal analysis PCA was used for data visualization in human gait investigation [156], to study motor unit discharge rates [115], to improve force estimation [145, 144] and, finally, PCA was used to reduce feature space dimensionality while keeping classification accuracy high [92, 47, 46] and improving the delay-accuracy trade-off [69, 25].

**3.3.2.1.1 Algorithm implementation details**   For PCA calculation we use the implementation of the Statistics Toolbox of MATLAB.

### 3.3.2.2 Independent component analysis

ICA is another unsupervised machine learning approach to find a linear transformation such that the resulting features are statistically independent. Other than PCA this holds true also for higher order dependencies. A classical example to explain the concept of ICA is the so called cocktail party problem. A person attending a cocktail party perceives a superposition of voices and other acoustic signals. To follow a discussion within such a noisy environment this person needs to be able to process the superposition such that it extracts the relevant signal, i.e. the voice of the discussion partner[3]. This is what is called the independent component, namely the signal stemming from a source which is statistically independent from the other sources contributing to the superimposed signal. More formally we would like to find the $K$ independent source signals $\{s_i\}_{i=1}^{K}$ from $N$ observations. We write these in a matrix $S$ with dimensions $K \times N$ and assume linear mixing by $A$ of dimensions $M \times K$ and the observations of the mixing of independent sources is denoted by $X$ and is of dimensions $M \times N$. Hence,

$$X = AS. \tag{3.1}$$

An ICA algorithms aim is to estimate $W \cong A^{-1}$ based on the data matrix $X$ where equality would be the ideal case.

There are several different algorithms for the estimation of the unmixing matrix $W$. In this work we use two very well known algorithms, namely, the information maximization approach to ICA [6, 2] and the FastICA algorithm [79]. While the former is using as objective function the entropy of the estimated independent components and, hence, takes into account all moments of the probability distribution, the latter is based on extremizing the kurtosis of the estimated independent components. This means that in a strict sense it can be only an approximation to the estimation of independent components unless it is assured that no higher order dependencies are present. However, FastICA is computationally much more efficient [79] than gradient descent (including the natural gradient descent [2]) algorithm used for the information maximization approach, hence the name. However, the idea to maximize or minimize the kurtosis is based on the assumption that the marginal distribution of the independent sources are non-Gaussian [79]. If we assume the independent sources to be the muscles, then non-Gaussianity seems a valid assumption given that not even their superimposed activity (cross-talk) as measured by the surface electrodes results in a Gaussian distribution. Nevertheless, in what follows we present a blind source separation technique that allows signals to be Gaussian and exploits instead their temporal structure.

### 3.3.2.2.1 Algorithm implementation details
Our implementation of the InfoMax algorithm is based on [7]. The number of sweeps is 30. Initially the learning rate is set

---

[3]The cocktail party is a nice metaphor to explain the problem with which ICA algorithms deal. However, how a brain actually solves this problem may be entirely unrelated to current ICA algorithms and is a prevailing topic of scientific research [160].

to $\gamma = 10^{-5}$, after 25 sweeps it is reduced to $\gamma = 10^{-6}$ for the remaining 5 sweeps. The weight matrix is initialized to the identity matrix.

We use the implementation of FastICA as provided online by [63]. We kept the recommended default values, i.e. a cubic non-linearity, unit step size, and no stabilization. To control convergence $\varepsilon = 10^{-4}$ and the maximum number of iterations is 1000. The weight matrix is initialized by random values uniformly distributed between 0 and 1.

### 3.3.2.3 Second order blind identification

The algorithm for second order blind identification (SOBI) was first proposed in [152] and further extended in [8, 9]. In this approach the independent sources are assumed to have a Gaussian distribution with some auto-correlation function different from the Dirac delta, i.e. they have temporal structure. The original algorithm termed AMUSE [152] is based on the idea to use a time-lagged covariance matrix to estimate the unmixing matrix $W$ [79]. The extensions include several time lagged covariance matrices [8, 9] and show an improved performance as well as a reduced dependence on the specific choice of the time lag $\tau$.

**3.3.2.3.1 Algorithm implementation details** We use the implementation of SOBI based on multiple time-lagged covariance matrices as provided online [10] and leave the default value as proposed by the authors for the number of estimated covariance matrices at 100.

### 3.3.2.4 Common spatial patterns

CSP is a supervised extension of PCA first proposed for the study of electroencephalogram (EEG) [95]. The CSP algorithm computes spatial filters that maximize the ratio of the variance of the data conditioned on one class to the variance of the data conditioned on the other class. In this way, spatial filters can be designed to extract those components of the data that differ maximally (in terms of the variance) between conditions [95].

The two class CSP algorithm then solves the optimization problem

$$\vec{w}^* = \operatorname*{argmax}_{\vec{w} \in \mathbb{R}^N} \left( \frac{\vec{w}^{\mathrm{T}} \Sigma_1 \vec{w}}{\vec{w}^{\mathrm{T}} \Sigma_2 \vec{w}} \right) \tag{3.2}$$

with $\Sigma_1$ and $\Sigma_2$ being the covariance matrices of the signals $\vec{x}$ collected during class 1 and class 2, respectively. $N$ is the dimensionality of the signal space. The solution can be obtained by solving the generalized eigenvalue problem

$$\Sigma_1 \vec{w} = \lambda \Sigma_2 \vec{w} \tag{3.3}$$

where $\lambda$ is the eigenvalue to the eigenvector $\vec{w}$. Preprocessing is then usually done by combining all eigenvectors of equation (3.3) to form the filter matrix $W$.

There are three common approaches to extend the two class algorithm to many classes: one-vs-one (OvO), one-vs-rest (OvR), and joint block diagonalization [65, 67]. OvO is a simple straight forward extension where one computes all $N_C(N_C - 1)/2$ pairwise combinations of classes. OvR in turn is defined by the following objective function

$$\vec{w_c}^* = \underset{\vec{w_c} \in \mathbb{R}^N}{\mathrm{argmax}} \left( \frac{\vec{w}^{\mathrm{T}} \Sigma_c \vec{w}}{\vec{w}^{\mathrm{T}} \left( \sum_{i \neq c} \Sigma_i \right) \vec{w}} \right) \tag{3.4}$$

In the denominator we sum over all covariance matrices that are conditioned on classes not being the currently chosen class $c$. Hence, this will give us a filter matrix $W \in \mathbb{R}^{N \times L \cdot N_C}$.

Hahne, et al. [67] show that the performance difference between OvO and OvR is statistically not significant while both methods show a slightly better mean performance than the joint block diagonalization algorithm. Interestingly, in [65] the authors prove analytically that the joint block diagonalization algorithm is equivalent to (unsupervised) ICA.

**3.3.2.4.1 Algorithm implementation details** For our investigation we implement the OvR algorithm and take for each class 30 eigenvectors, i.e. filters or components having the largest eigenvalues. Thus we get $30 \cdot 8 = 240$ spatial filters.

### 3.3.2.5 Results of spatial filtering with data-dependent filters

Difference between the two ICA algorithms is negligible. However, given that we need one algorithm to compare to the other filters we take ICA based on information maximization as its mean performance is sometimes slightly better as can be seen in figure 3.7.

## 3.4 Feature selection

Feature selection is used in this thesis to address the following questions: 1.) "How many channels can be added, and still provide an improvement in control accuracy?". This question was asked by Parker, et al. in [108, Chapter 18]. Yet no systematical study was performed to address this question. 2.) What are "good" positions for electrode placement? Clearly, the location where the electrode is put on the forearm influences classification accuracy as the signal amplitude drops quickly with the distance from a muscle. To address this question and to propose a standardization for sensor placements to improve reproducibility of sEMG studies, the European concerted action SENIAM (surface EMG for a non-invasive assessment of muscles) addressed the issue of sensor placement [70]. Their suggestion is to place electrodes on the muscle belly but not above the innervation zone (sometimes also called motor point). Signals above the innervation zone are "not typical" for the muscle, since small electrode displacements result in large
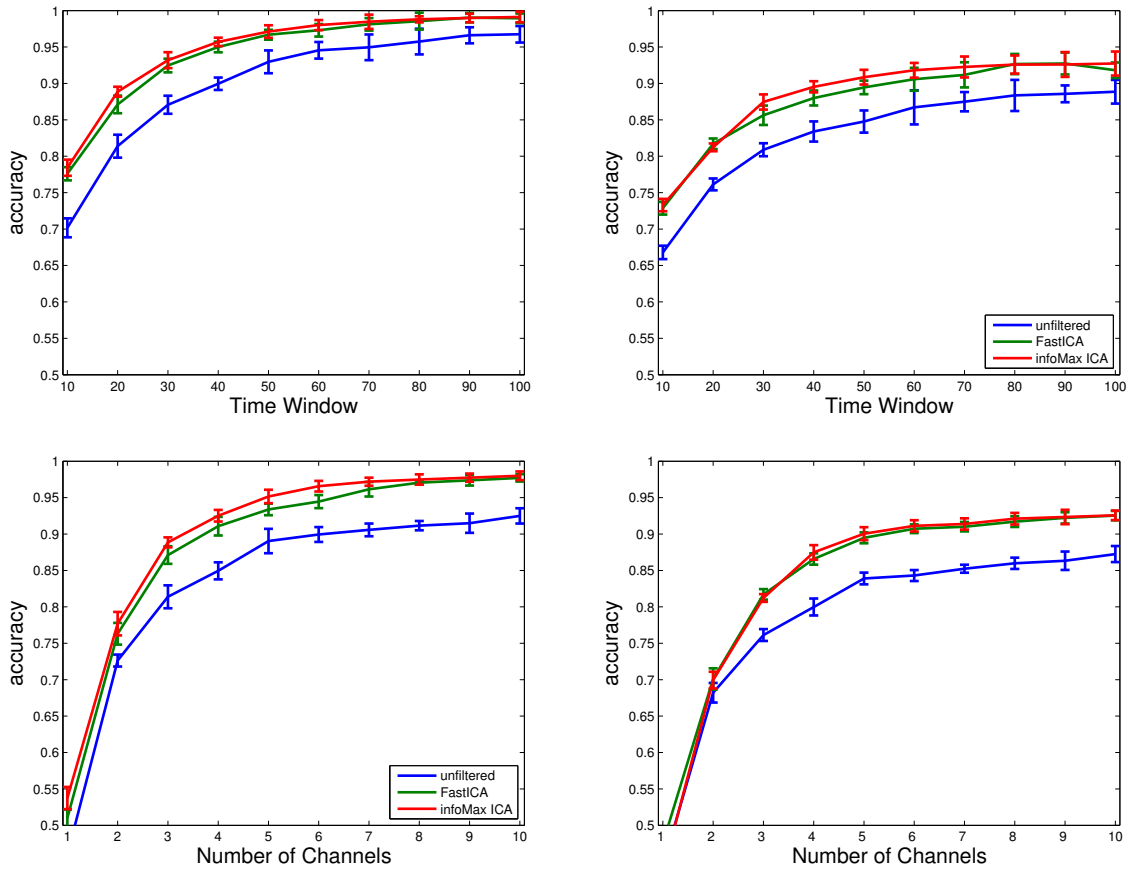
Figure 3.7: Comparison of ICA algorithms with unfiltered classification accuracy. **Upper** graphs: Accuracy over time window length, classification is done with the three best channels. **Lower** graphs: Accuracy over number of channels, time window is set to 20 ms. **Left** is static regime only while **right** figures result from including the dynamic regime.

Figure 3.8: Comparison of classification performance employing data driven spatial filters and longitudinal bipolar with the unfiltered case. **Upper** graphs: Accuracy over time window length, classification is done with the three best channels. **Lower** graphs: Accuracy over number of channels, time window is set to 20 ms. **Left** is static regime only while **right** figures result from including the dynamic regime.
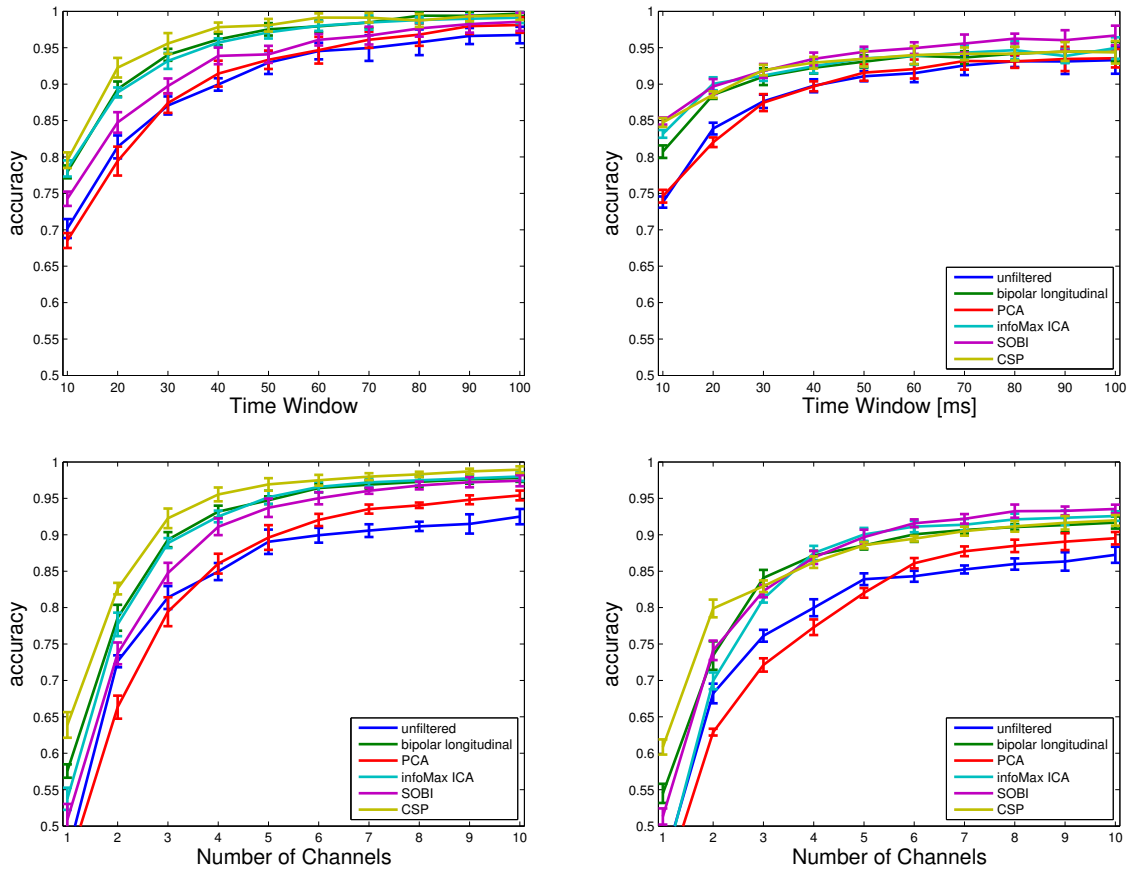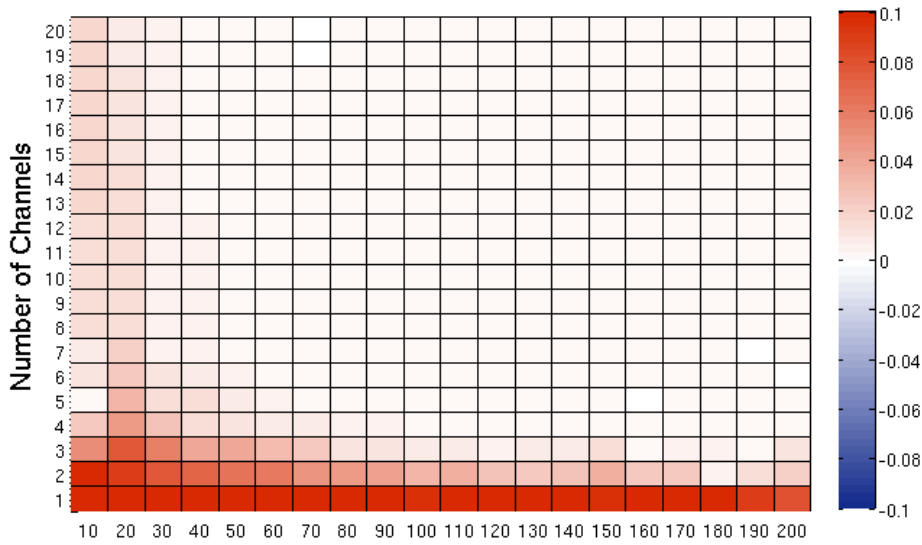
(a) Static regime only.
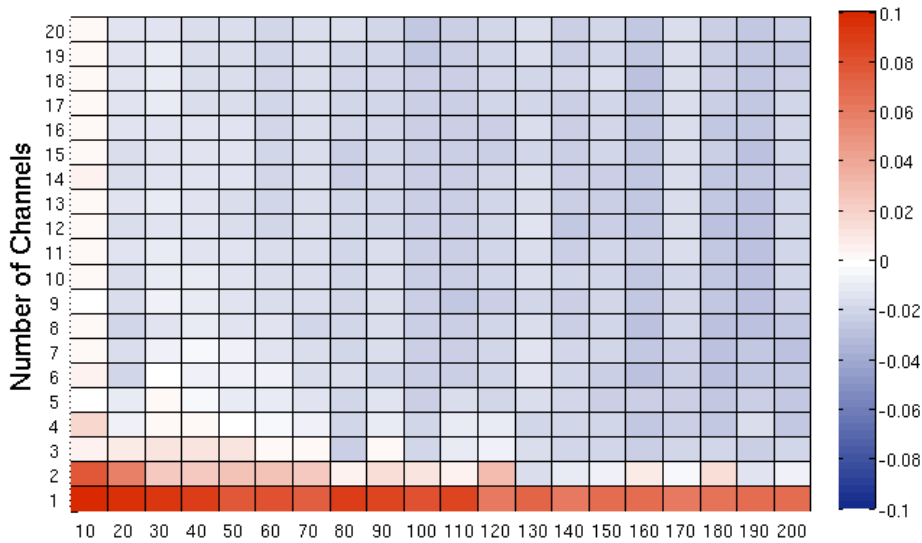


(b) Static and dynamic regime.

Figure 3.9: Direct comparison of CSP and SOBI for signals measured during static contraction (**a**) and signals measured static contraction and dynamic contraction onset and offset (**b**). Color plots of difference of classification accuracies after filtering with CSP and SOBI. Positive (red) values mean CSP filtering is better while negative (blue) values mean SOBI filtering results in a better classification performance.
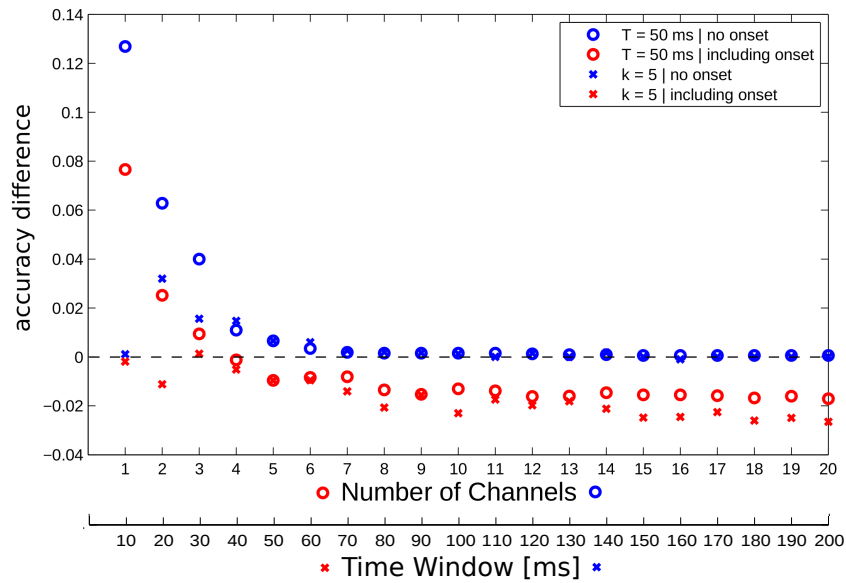
Figure 3.10: Excerpt of comparison between CSP and SOBI figure 3.9. We plot here column five and row five of the color plots in figure 3.9. Circles mean columns of the color plots above (fixed time window $T = 50$ ms), i.e. the respective x-axis is "Number of Channels" and crosses the rows (fixed number of channels $k = 5$), i.e. the x-axis here is "Time Window". Blue marks static regime only while red marks results from static and dynamic regime.

signal amplitude variations. Sensors should preferably be placed in the middle between muscle end zone (tendons) and innervation zone and away from other muscles to avoid cross-talk [70].

As we measure with two high density electrode arrays surrounding the entire forearm we pick up signal from all places mentioned, i.e. muscle belly, innervation zone, close to tendons and some electrodes' position is between muscles, hence, crass-talk will be strong in those channels. In order to assess which position is a "good" position and answer question 2.) we will employ feature selection algorithms to select a subset $k$ out of the 126 signals. Furthermore, by varying $k$ and assessing classification accuracy as standard offline measure of control accuracy we can address question 1.).

In principle, to find the best electrode set is an NP complete [62] problem therefore we have to resort on heuristics. An exhaustive search would amount to evaluate $2^{126}$ channel combinations. Even if we restrict the number of channels, i.e. we look for the best subset of $k$ channels we would have to evaluate $\begin{pmatrix} 126 \\ k \end{pmatrix}$ which grows quickly with $k$. As we do not know in advance which heuristic is suited best to our problem, we perform a comparative study and test feature selection algorithms from each of the three algorithm categories: filters, wrappers and embedded methods [66].

In the following we give a quick overview of feature selection and electrode positioning performed within the EMG community. Kendell et al. [89] investigate electrode pair selection and positioning based on signal power and frequency characteristics. The goal in their study is to find a good electrode placement for use in clinical rehabilitation, hence their objective is not to improve pattern recognition accuracy. Their approach takes into consideration two aspects of the signal: the amplitude should be high and the mean and median power spectral frequency should be small. A large amplitude provides a good signal-to-noise ratio and the latter index gives an electrode positioning far from the innervation zone (IZ) which is known to lead to poor signals [70].

Hargrove et al. [69] apply sequential backward selection after a high dimensional feature space is produced by PCA preprocessing. The aim is to reduce dimensionality for saving computation time while keeping a good level of classification accuracy. Pulliam et al. [132] apply a sequential forward selection to select the best muscles for optimizing the performance of a time delayed artificial neural network for the control of a transhumeral prosthesis[4]. Sequential forward selection was also applied in [114] to find the best features for classification among 19 auto-cumulants of order two, three and four.

Next we present present briefly the different feature selection methods whose performance we will compare. Please note that despite the fact that electrodes, channels and features have different meaning in general it is correct to use these terms interchangeably in what follows as the selection of features, i.e. input dimensions to our classifier, is identical to select specific electrodes which is again identical to a given signal channel as

---

[4]Although one might argue that given only 7 features (in this case each feature corresponds to a muscle due to specific electrode placements) the best choice would be an exhaustive search evaluating all 128 possibilities.

each channel is the root mean square of a monopolar electrode signal. However, in general a channel can be different from an electrode as in the case of blind source separation techniques where a channel denotes a source, i.e. a weighted sum of electrode signals. There is no clear definition of a channel in the myoelectric community and it shall not be of our concern in this study to provide one. Features is an even more general term as it stands for the input dimensions to a classifier which can be any function of the raw signal as measured from an electrode.

We categorize the feature selection approaches into wrapper methods, which are search strategies that have as objective function the classification accuracy of the classifier that is employed for classification afterwards, that is, these search strategies optimize the quantity that is our general assessment quantity for good electrode positions. In contrast to this, filter methods by definition have an objective function different from the classification accuracy. We will study information theoretic quantities as objective function. However, the final assessment of the quality of selected channels still remains the classification accuracy of the LDA classifier. Embedded methods in turn are classifiers where the feature selection is intrinsically part of the algorithm. In our case random forests [15] and AdaBoost [60]. The selection of channels done by those algorithms are then again assessed in terms of classification accuracy of an LDA.

Apart from this general categorization we have to differentiate between two conceptually distinct search heuristics within the wrapper methods, namely deterministic and stochastic selection methods as is seen next.

## 3.4.1 Wrapper methods

### 3.4.1.1 Deterministic search heuristics

**3.4.1.1.1 Sequential forward selection.** Together with backward selection (also called backward extraction) this approach is the simplest possible. The best feature subset $\mathcal{S}$ according to this algorithm is selected sequentially by first evaluating for each feature alone the classification accuracy and taking the one with the highest score. Next all pairs are evaluated with the one selected and those remaining and then the best pair is chosen to be $\mathcal{S}$. This procedure is iterated until $\mathcal{S}$ has a cardinality of $k$, the predefined subset of best features size.

**3.4.1.1.2 Sequential forward floating selection.** It is known, however, that sequential forward selection (SFS) has the deficiency of getting easily stuck in local minima. In the feature selection context, this is also called the "nesting effect" [66, Chap. 4]. This deficiency is overcome by many different extensions of the SFS heuristic [66, Chap. 4]. Here we will test the extension called sequential forward floating selection (SFFS) which was first proposed in [131].[5] It is parameter free, simple and it was shown empirically to produce good results [66].

---

[5]Please note that there was an error in the original article, a correct version can be found here [66]

The general idea is to not only perform a forward selection but also a backward extraction if the resulting feature set provides a better performance then the feature set with the same cardinality tested before. In such a case forward selection would proceed with a different (better) set of features as would be by only forward selecting features.

### 3.4.1.2 Stochastic search heuristics

**3.4.1.2.1 Particle swarm optimization.** The particle swarm optimization (PSO) algorithm was first proposed in [90]. Since then many different variations of the original idea were proposed [128]. However, the nowadays most widely spread version of PSO, the so called canonical PSO [128], is the choice for our study.

PSO consists of a swarm of particles, each particle is a distinct subset of selected channels with cardinality $k$ for which the accuracy is evaluated. The idea is that these particles explore the landscape defined by the objective function to find the optimum.

**3.4.1.2.2 Discrete particle swarm optimization.** Given that our problem is discrete in its nature (the search space is discrete, actually binary: either a feature is selected or not) we use the adapted version of discrete PSO [91] proposed for feature selection in [1]. As the authors have shown this PSO algorithm outperforms also simulated annealing. Note: in the review [128], formula (6) is wrong. The logistic function is a function of the velocity not the position of the particle, confront [91].

## 3.4.2 Filter methods

Filter approaches to feature selection are based on objective functions other than the objective function of the learning algorithm. Here we compare four different information theoretic objective functions:

- mutual information feature selection (mifs) [5]

$$J_{\mathrm{mifs}}(X_k) = I(X_k; Y) - \beta \sum_{X_j \in \mathcal{S}} I(X_k; X_j) \tag{3.5}$$

- maximum relevance minimum redundancy (mrmr) [127]

$$J_{\mathrm{mrmr}}(X_k) = I(X_k; Y) - \frac{1}{|\mathcal{S}|} \sum_{j \in \mathcal{S}} I(X_k; X_j) \tag{3.6}$$

- joint mutual information (jmi) [158]

$$J_{\mathrm{jmi}}(X_k) = \sum_{j \in \mathcal{S}} I(X_k X_j; Y) \tag{3.7}$$

- conditional mutual information maximization (cmim) [56]

$$J_{\text{cmim}}(X_k) = \min_{X_j \in \mathcal{S}} I(X_k; Y | X_j) \qquad (3.8)$$

All these objective functions can be derived as approximations to the maximum of the conditional likelihood $\mathcal{L}(y|\vec{x}, \theta)$ of the labels $y$ given a feature vector $\vec{x}$ and some parameters $\theta$, as is elegantly proved in [18]. For an explanation of the information theoretic quantities please refer to A. The estimation of the information theoretic quantities is based on histogram estimations and $\beta = 1$ in equation (3.5) as proposed in [18].

While the equations above define the objective functions, the selection algorithm is deliberately chosen to be the simplest one, namely sequential forward selection.

According to [17, 18] the best performance is given by $J_{\text{jmi}}$. In 3.4.4.2 a comparison of these objective functions in terms of classification accuracy within our standard pattern recognition setting (which is defined earlier) can be found.

## 3.4.3 Embedded methods

### 3.4.3.1 AdaBoost

AdaBoost[6] is a well studied ensemble classifier first proposed in [59] that relies on the concept of boosting. The motivation for boosting is a procedure that combines the outputs of many "weak" classifiers to produce a powerful "committee". The weak classifier used here is a classification and regression tree [60], one is trained for each feature. Features are then selected by the weights that the AdaBoost algorithm gives the single classifiers and in that sense the individual features.

### 3.4.3.2 Random forests

Random forests is another example of an ensemble classifier [15]. Also here feature selection is intrinsic to the algorithm and therefore counts to embedded methods [66, Chap. 5]. The selection is performed by an internal variable importance estimate which gives a natural ranking for the features on which a random forest is grown [15]. We refer the reader to the original publication for details about the ranking measure [15]. Here we used the implementation of random forests as can be found in [81]. There are two parameters to be set: The number of trees we set to 500 and the number of features per tree is set to one third of the total number of features.

## 3.4.4 Comparing performances of feature selection algorithms

### 3.4.4.1 Comparing search strategies

Which search strategy finds the best channels? Here we compare the two deterministic with the two stochastic search or optimization methods presented in 3.4.1. As can be

---

[6]Sometimes also called AdaBoost.M1 [60]

seen from figure 3.11, the superiority of SFFS over SFS for myoelectric channel selection is very small. The algorithms are most of the times equal in performance, sometimes SFFS is slightly better (less than 1%) and only for a time window of 10 ms SFFS proves to improve performance more than 2% over SFS.


Comparing PSO with discrete particle swarm optimization (DPSO) in figure 3.12 shows that the performance is virtually the same: Sometimes PSO and other times DPSO performs slightly better (difference in accuracy < 1%) while most of the times their performance is equal. However, as DPSO is suited better to our problem due to its discrete nature we choose this algorithm for the general comparison in 3.4.4.3.



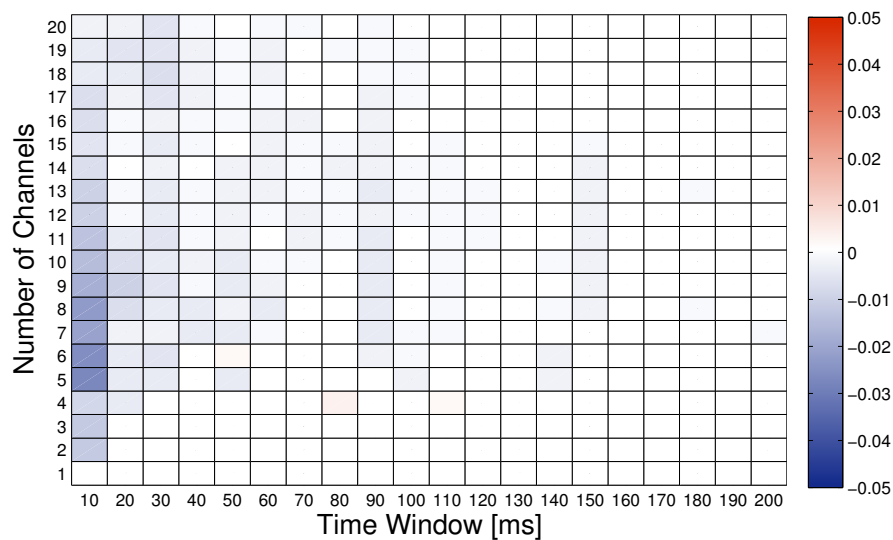Figure 3.11: Difference of classification accuracy as realized by the deterministic search strategies SFS and SFFS. Negative (blue) values denote SFFS found the better channels while positive (red) values denote that SFS found the better channels.
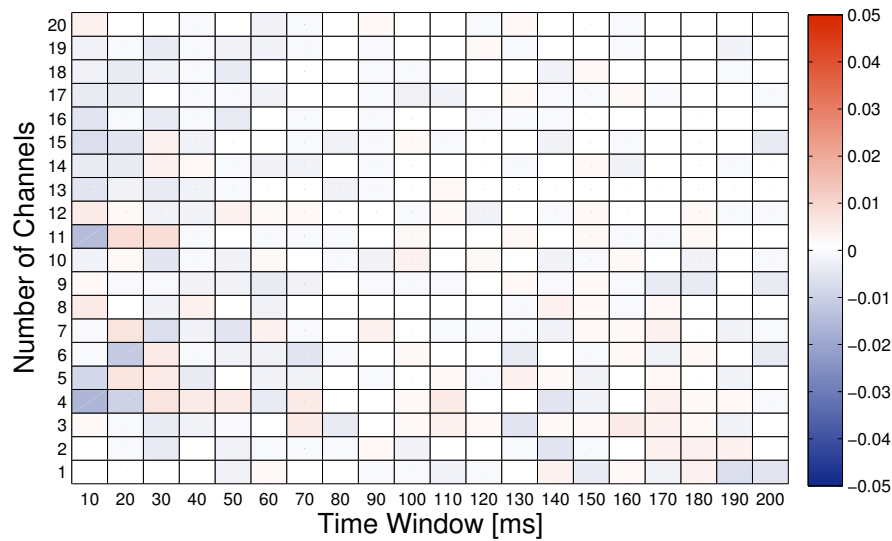
Figure 3.12: Difference of classification accuracy as realized by the probabilistic search strategies PSO and DPSO. Negative (blue) values denote DPSO found the better channels while positive (red) values denote that PSO found the better channels.

### 3.4.4.2 Comparing information-theoretic objective functions

Which information theoretic objective function gives the best result for the filter methods (see 3.4.2)? We compare the objective functions defined by the equations (3.5), (3.6), (3.7) and (3.8). The search strategy is SFS and accuracy is computed by the standard classifier scheme as defined earlier based on the respective selections.

We find that jmi gives the best results. Interestingly, this is in accordance with tests on other data sets [18]. Second best objective function is cmim. This objective function outperforms jmi only a few times by less than 2% as can be seen in graph d) of figure 3.13. The mrmr and mifs perform poorly in comparison to the other two especially for a channel number smaller than nine.

### 3.4.4.3 Comparison of feature-selection methods

## 3.4.5 Placement of electrodes

We will now turn to the question of which electrode positions provide good performances in terms of classification accuracy.

In order to approach the question whether specific electrode positions provide good performance quality or whether there are rather good areas for electrode positioning, we select a set of electrodes and perform an exhaustive search of the positions in their
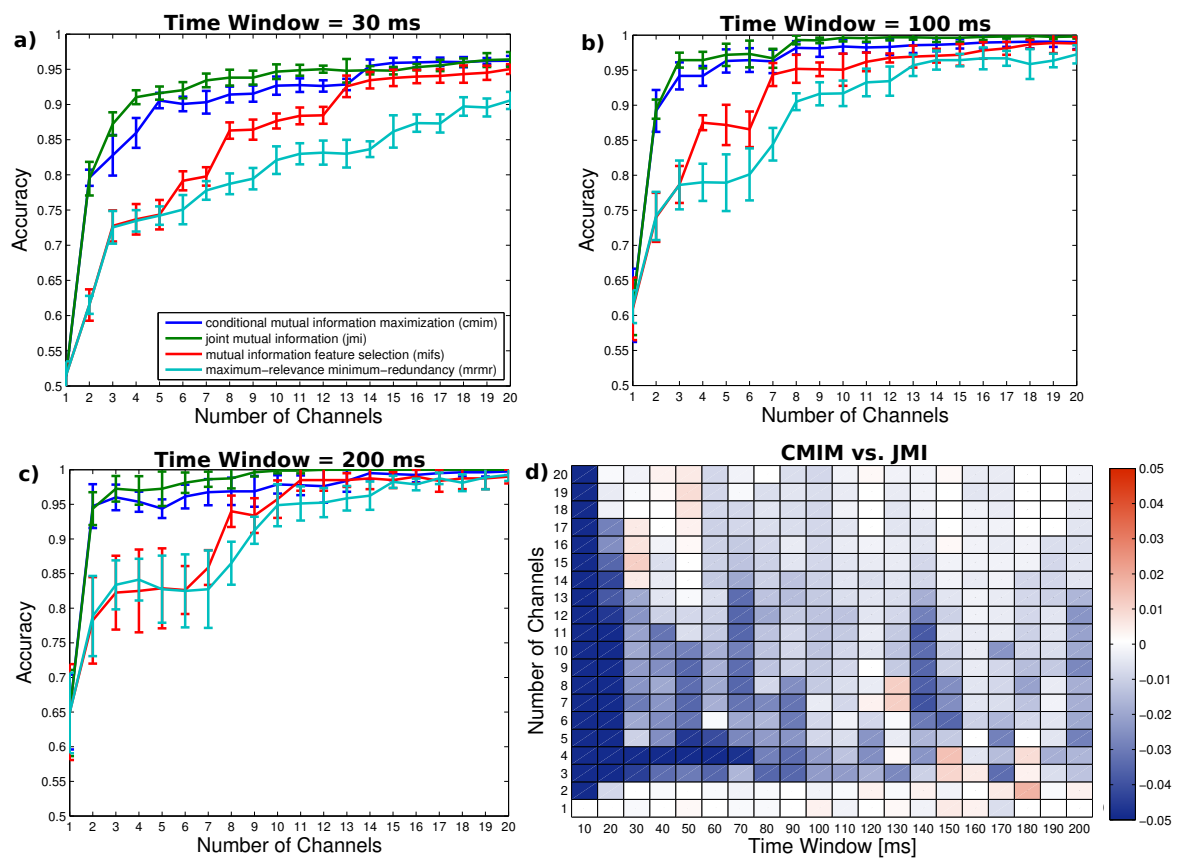
Figure 3.13: Quality of channel selection by filter methods using information theoretic objective functions.
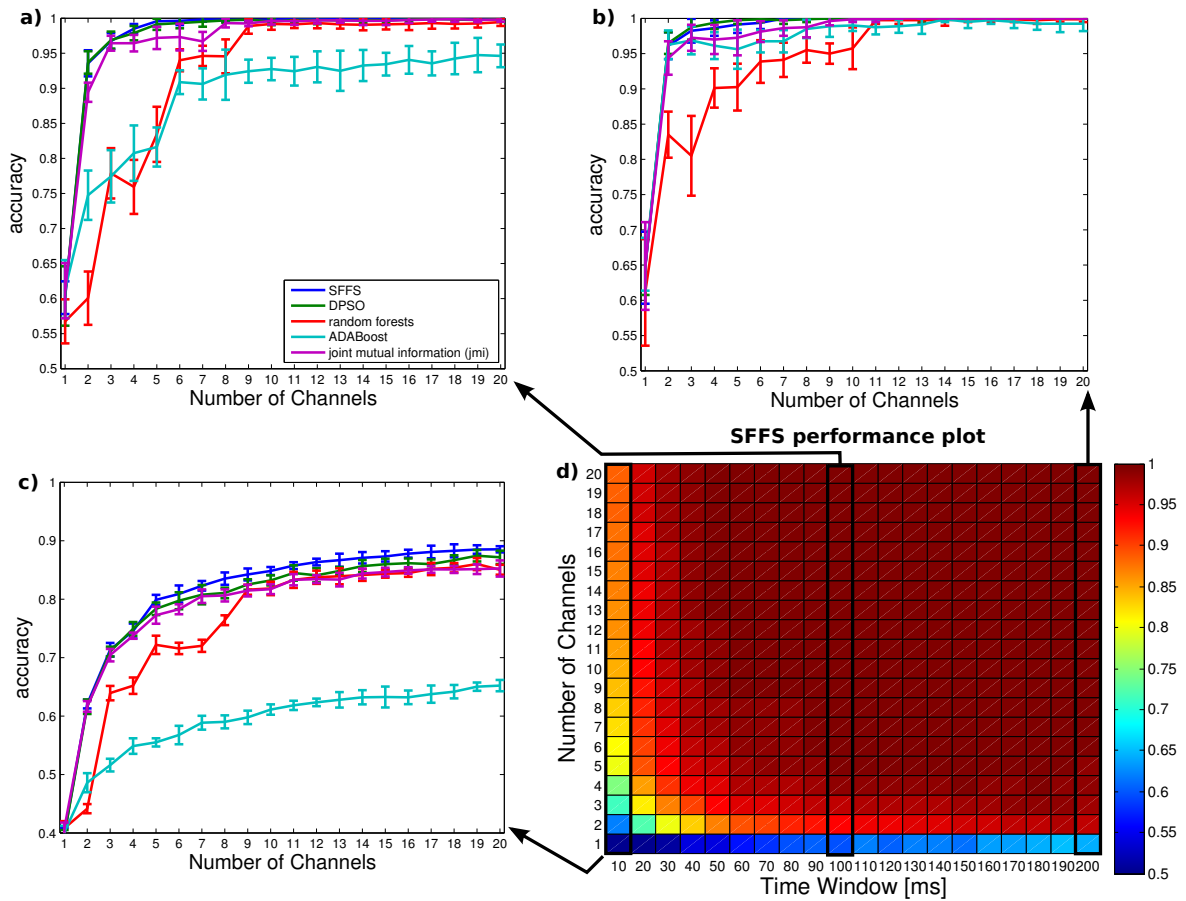
Figure 3.14: Comparison of different feature selection algorithms for different time window lengths as indicated by the arrows from the color plot, **a)** 100 ms, **b)** 200 ms and **c)** is for a time window of 10 ms. Color plot shows the trade-off between number of channels and time window length when channel selection is performed by the best algorithm, SFFS.
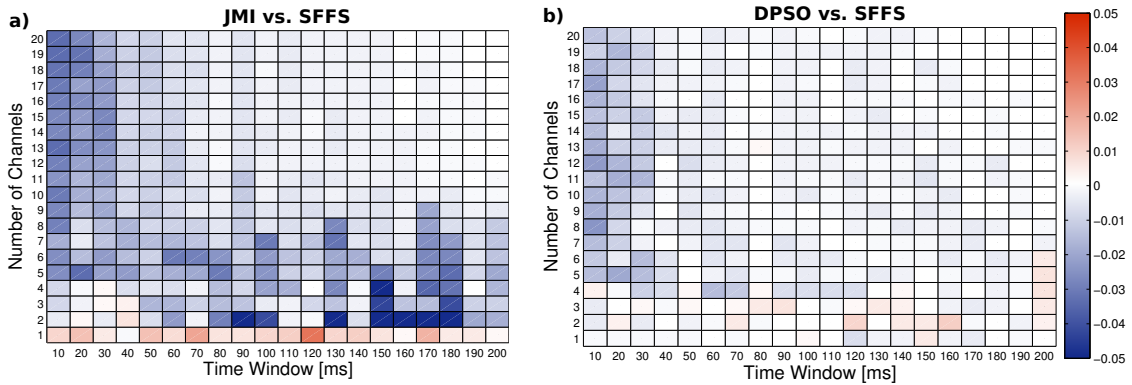
Figure 3.15: Color plot of accuracy difference to directly compare DPSO and jmi to SFFS. Negative values (blue) mean SFFS is better while positive values (red) mean the respective other algorithm is better.
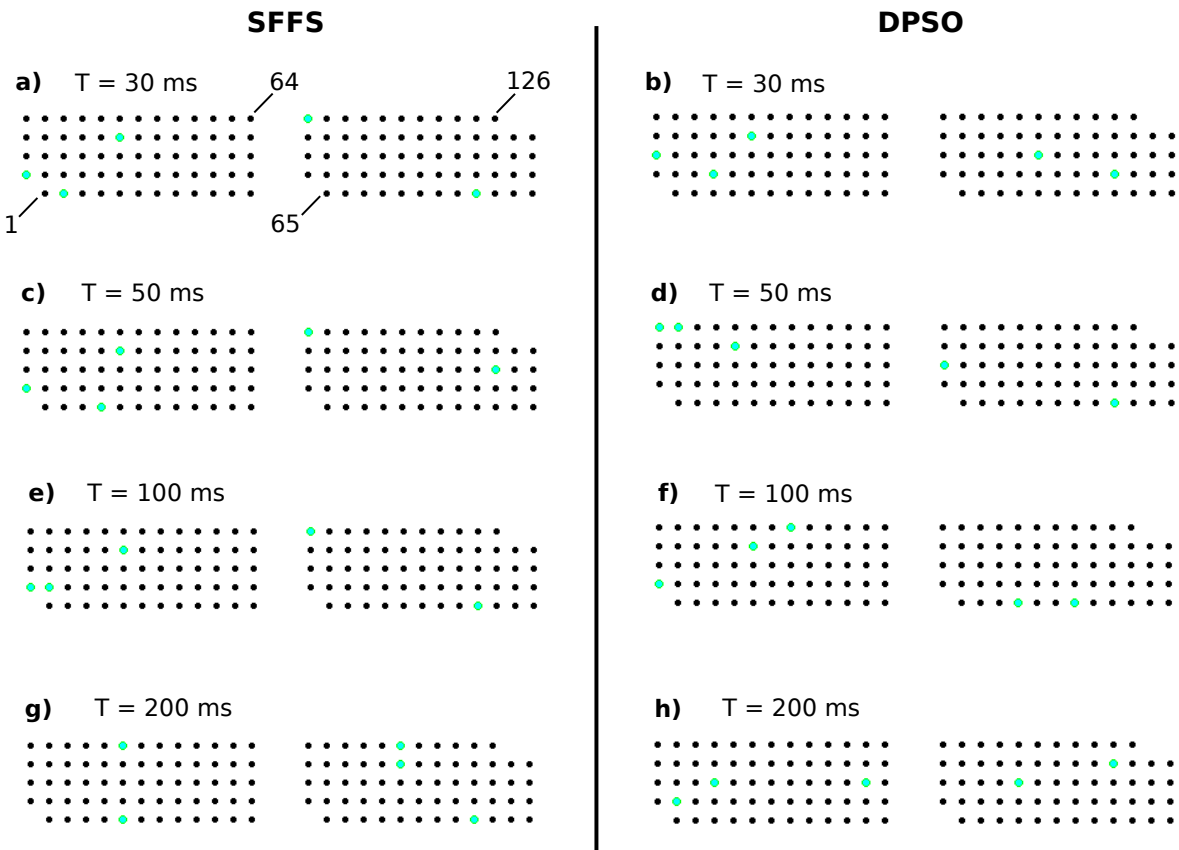


Figure 3.16: Five best electrode positions as selected by SFFS and DPSO for different time window lengths.

neighborhood. Considering only neighbors along rows and columns of the electrode grid, the maximal number of possible electrode configurations that need be tested are $5^k$ where $k$ is the number of preselected electrodes. Note that electrodes positioned at the top or bottom (long) array edges lack, respectively, the upper or lower neighbor, while electrodes positioned at the lateral (short) edges of the electrode array are assigned neighboring electrodes from the respective other end of the respective other array.
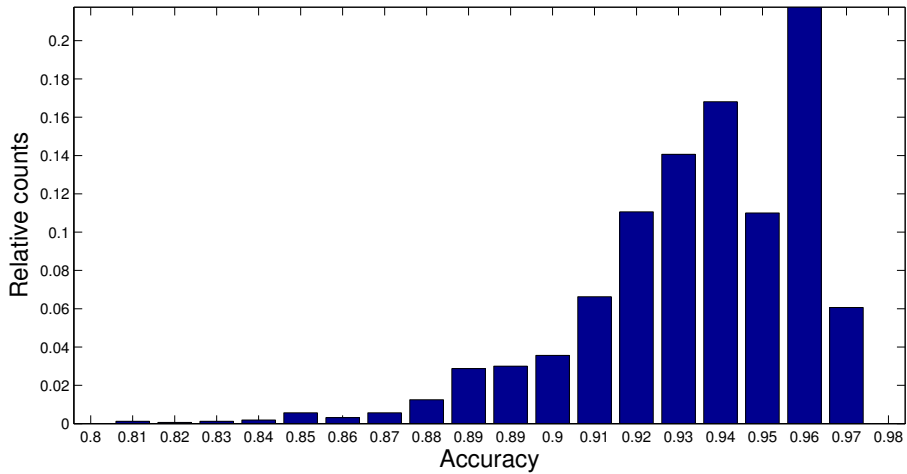
(a) SFFS selection of primary electrode configuration.
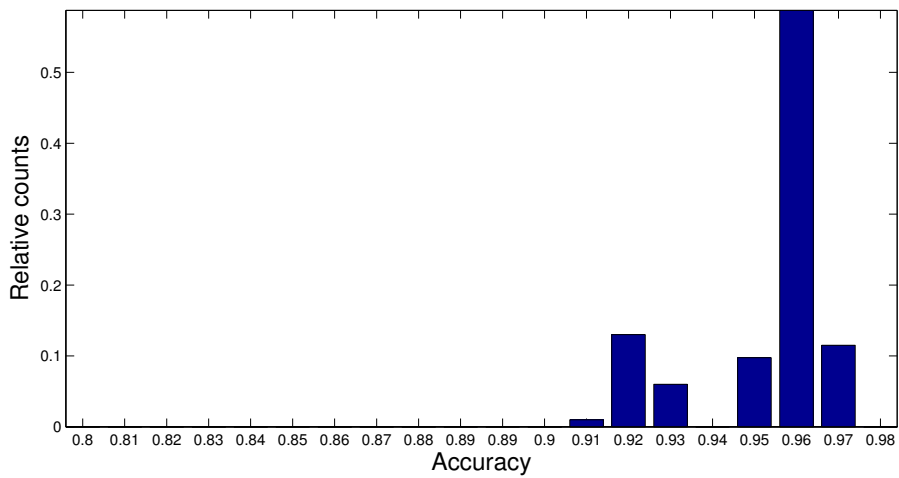


(b) DPSO selection of primary electrode configuration.

Figure 3.17: Histograms of classification accuracy when classifying with all possible combinations of the neighborhood and $k = 5$ selected channels. (a) selection performed by SFFS, (b) selection performed by DPSO. In both situations $T = 50$ ms which is one of the cases where their performance is equal (see figure 3.15), namely, 96.6%.

# 4 Discussion

## 4.1 Probability density function

The probability density is the central quantity of stochastic approaches to data analysis. With knowledge of the probability density function from which a data set is drawn, one can in principle derive a Bayes optimal classifier or find parameter-based estimates of information-theoretic quantities, which is especially important in the case of high dimensional data.

It is often assumed that myoelectric signal statistics follow a Gaussian distribution, however, some investigators found that the distribution of myoelectric signals is stronger peaked than predicted by a Gaussian probability density function [109, 78, 12, 33]. Clancy and Hogan [33] proposed the Laplace distribution as an alternative statistical description of MES. They concluded that the true distribution lies "in between" [33] a Gaussian and a Laplace distribution.

We further investigated the functional form of the probability distribution and found that many of the 126 available signals at each measurement are clearly asymmetric, as can be seen by visual inspection as well as quantified by the third moment of the probability distribution, known as skewness. However, the Gaussian as well as the proposed Laplace distribution are symmetric and hence cannot provide a correct description of myoelectric signals. Furthermore, for those signals with a small skewness, i.e. with a symmetric distribution, we found deviations from a Gaussian. These deviations occurred not only at the peak of the distribution but also at its tails. The tails often show a linear rather than a quadratic decay as implied by the Gaussian model, which becomes obvious when plotted in a semi-logarithmic plot. Hence the behavior is closer to a Laplace distribution. Still, the fit of this distribution to the data is not adequate when estimating the parameters of the Laplace density. We, therefore, proposed a mixture model of a Gaussian and a Laplace random variable with mixture parameters obtained by the best fit in terms of KL-distance. This mixture then captures the peak as well as the tails of the distribution which can also be appreciated by visual inspection.

In the following we are discussing a symmetric mixture distribution, although this may not explain part of the myoelectric signal statistics. A more complex model could be based on a generalized asymmetric form of the Laplace distribution [96] or be obtained by considering the positive and the negative signals separately. We have chosen not to follow this path because of the possibility of an analytical treatment of Bayes classifiers and Bayesian filtering does not seem to be feasible for a more complex form of the model

distribution.

## 4.2 Amplitude estimation

Amplitude estimation is of general importance for clinical applications of electromyograms. The maximum likelihood estimator of the myoelectric signal amplitude depends on the likelihood model, i.e. on the probability density function that supposedly underlies the given data set. For a centered Gaussian distribution the RMS provides a sufficient statistics, while for a Laplace distribution the mean absolute value (MAV) can be used, as discussed in [33]. These authors found that the signal-to-noise ratio for MAV amplitude estimation is 2.0%-6.5% higher than RMS estimation [33].

In order to improve the signal-to-noise ratio we studied an advanced amplitude estimation technique from the family of Bayesian filters proposed in [138]. We showed that the Bayes filter produces a 4.3-fold better signal-to-noise ratio when a Laplace likelihood function is chosen as compared to RMS estimation with a 200 ms smoothing window. Furthermore, when applied to artificial data the Bayesian filter produces a four-fold smaller root mean square error than the RMS with a time window of 250 ms. Using longer time windows the estimation with RMS will improve but also increases the delay of estimation of transient signals, which, depending on the specific application, might not be tolerable. In application to prosthetic control, for instance, delays should not exceed 300 ms as longer controller delays are perceivable by the user [54, 45].

We conclude that off-line tests show very promising results of the algorithm for enhancement of proportional and simultaneous control of multiple degrees of freedom of a powered prosthesis. Nevertheless, the final quality assessment in a real time on-line test is still in process and must be completed before it can be implemented as part of the control algorithm of a myoelectrically controlled prosthesis.

Finally, we extended the original Bayes filter proposed in [138] by introducing an inverse Gamma distribution as model for the prior. In doing so and by the approximation of $\sigma_{\mathrm{MVC}} \to \infty$, we derived explicit equations that describe parameters updates of the inverse Gamma distribution for the evolution and observation update. The performance improvement for this extension may be marginal. However, it is a first step towards a multi-dimensional Bayes-Fokker-Planck filter which can be realized by replacing the inverse Gamma distribution by its multi-dimensional counterpart that is known as the inverse Wishart distribution [118].

## 4.3 Pattern recognition

### 4.3.1 Spatial filtering

Spatial filtering improves muscular unit action potential extraction [106] and also the classification accuracy of pattern recognition systems [75, 67]. In a pioneering study

Huang et al. [75] found that of a set of different data independent spatial filters the double differential gave the best result for classification improvement. However, we compared the proposed data independent filters and found that the simple bipolar filter performed best if aligned with muscle fibers. This finding is important because it gives additional justification to the widely spread use of bipolar electrodes in clinical applications [108, chapter 5]. Furthermore it suggests that a bipolar filter aligned with the muscle fibers is a good reference to evaluate performances of other spatial filters.

Next we included several data driven filters in our comparative study and found that CSP performed best, thus supporting the findings of Hahne, et al. [67]. However, despite the many studies on using pattern recognition systems for myoelectric control, many of them (e.g. [67, 75]) ignore the dynamic regime at the transition from one contraction to the other. Including this regime is known to reduce classification accuracy considerably [103, 44, 71]. However, high usability requires the controller to produce a robust classification also during dynamic contractions. When we include the dynamic signal regime, we find that another algorithm, SOBI, outperforms all other approaches and gives an improvement with respect to CSP by about 1-3% for time window lengths longer than 10 ms and more than four channels.

The improvement in performance of SOBI when the dynamic regime is included is explained by the algorithm's strategy to do source separation. It is based on time-lagged covariance matrices and so exploits the temporal structure of the data that is not present in the static regime. Moreover, we should add that CSP is supervised while SOBI is unsupervised. This makes SOBI applicable in situations where data is not segmented into classes, as is the case in, for example, regression approaches.

In general, we find that spatial filtering makes the algorithm less susceptible to the trade-off between accuracy and delay as the performance improvement is considerable for short time windows.

## 4.3.2 Feature selection

In general the question of how many channels can be added and still provide an improvement of control accuracy [108] is dependent on the specific pattern recognition system, the number and type of classes, on the channel positions and the measurement hardware. Our setup consists of two high-density electrode arrays and, hence, our spatial sampling frequency is high and, therefore, we do not need to worry about channel positions. Indeed, we picked a subset of best (w.r.t. a specific algorithm) electrodes out of the 126 available electrodes and thus gain information about "good" electrode positions.

Finding the best subset of 126 electrodes is an NP-hard optimization problem such that an exhaustive search becomes difficult. Therefore, it is preferable to use heuristics instead to find or approximate the optimal solution. Given that we cannot decide a priori which algorithm is suited best to our specific subset selection problem, we need to compare different algorithms and find that SFFS gives the best performance. This approach counts to the wrapper methods because the objective function is the classifica-

tion accuracy of an LDA classifier, and, hence, its selection will depend on the classifier used. To get classifier independent results we approach the subset selection problem with filter methods based on information-theoretic objective functions. From the four we tested the joint mutual information proved to be the best yielding results comparable to SFFS while being computationally much more efficient.

Turning back to our original question about how many channels can be added until performance is saturated (i.e. classification accuracy $> 99.9\%$) we sequentially increase the number of channels and vary the time window length of RMS calculation. We find that five channels suffice when a time window of 200 ms is used which is a common choice in other studies. However, if we instead prefer a shorter time window and still intend to obtain a comparable classification performance, we have to employ more channels. It seems, however, unlikely to reduce the time window below an interval of 50 ms where a practically relevant performance can still be achieved with 13 channels

For seven out of the eight situations considered here, three out of five electrodes are selected from the array that is fixed on the dorsal side of the upper forearm, i.e. above the extensor muscle group. Furthermore, we notice that most of the times electrodes selected on array one belong to the left half of the array, close to the ulna (elbow bone) and, hence, covering activity from extensor capri ulnaris, extensor digiti minimi, extensor digitorum and extensor carpi radialis brevis. The electrodes that are selected from the other array are located on its left and right part leaving the middle of the array empty. They are placed over the flexor carpi ulnaris and the flexor carpi radialis.

Finally, we addressed the question whether good positions for electrodes are single hot spots or larger areas. Performing an exhaustive search of the neighborhood of five preselected electrodes, we found that for the SFFS selection, accuracies can drop by about $10\%$ when choosing direct neighbors while for an electrode configuration selected by DPSO most of the configurations (close to $60\%$) fall into a range of about $\pm 0.5\%$ deviation from the accuracy that was produced by the original selection. However, also here we find a noticeable fraction of electrodes that cause a drop of the accuracy by about $5\%$. Hence, we conclude that there are areas which enhance robustness with respect to shifts of the electrode position, while in general position shifts can lead to significant performance losses.

# A Information Theory

## A.1 Basic concepts

While the term *information* (content) of myoelectric signals tends to get used in a colloquial way, in this thesis the term information will always denote Shannon information [142] which is quantifiable by the entropy

$$H(X) = -\sum_{i=1}^{N_E} P(X = x_i) \log_2 \left( P(X = x_i) \right) \tag{A.1}$$

$X$ is a discrete stochastic variable with realizations $x_i \in \mathcal{A}$, where $\mathcal{A}$ (often called the alphabet) is the set of possible events. The cardinality of $\mathcal{A}$ is denoted by $N_E = |\mathcal{A}|$. The basis of the logarithm defines the units in which information is measured. We will set it to 2 and thus quantify the information in bits. For a continuous stochastic variable with a probability density $p(x)$ the analogous quantity is called differential entropy [36]

$$h(x) = -\int_\Omega p(x) \log_2 \left( p(x) \right) \mathrm{d}x \tag{A.2}$$

where $\Omega$ is the support of the probability distribution $p(x)$.

Entropy quantifies information, namely the expected information that we obtain by observing the outcome of a stochastic variable, or, analogously, the expected uncertainty about the outcome before the observation.

Before we turn to our specific setting we need to define two further quantities: The conditional entropy $H(Y|X)$ and the mutual information $I(X, Y)$ [36]. The conditional entropy is defined by

$$H(Y|X) = E_{P(X,Y)} \left( \log_2 \left( P(y|x) \right) \right). \tag{A.3}$$

It describes the expected uncertainty left about the outcome of $Y$ given that we know the outcome of $X$. The conditional differential entropy is defined analogously.

Mutual information tells us how much information can be expected from the outcome of one stochastic variable about another. It is defined as

$$I(X, Y) = E_{P(X,Y)} \left( \log_2 \left( \frac{P(X,Y)}{P(X)P(Y)} \right) \right). \tag{A.4}$$

Note that the mutual information $I(X, Y) = D_{\mathrm{KL}} \left( P(X,Y) || P(X)P(Y) \right)$ can be expressed using the Kullback-Leibler divergence $D_{\mathrm{KL}}(p||q)$, see equation (2.10). As can

be seen from equation (A.4) $I(X, Y)$ is zero iff the stochastic variables are independent which implies that the joint distribution factorizes, and it is symmetric in its arguments. Furthermore, the mutual information can be expressed in terms of entropies

$$I(X, Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \tag{A.5}$$

Given that the conditional entropy is smaller or in case of independence equal to the marginal entropy and from the equality of the latter two entropy sums we can deduce that $I(X, Y) \leq \min(H(X), H(Y))$. Equality is exactly the case if one variable fully determines the other one (with lower entropy).

# Bibliography

[1] Dimitris K Agrafiotis and Walter Cedeño. Feature selection for structure-activity correlation using binary particle swarms. *Journal of medicinal chemistry*, 45(5):1098–107, February 2002.

[2] Shun-Ichi Amari, A. Cichocki, and H.H. Yang. A new learning algorithm for blind signal separation. *Advances in neural information processing systems*, pages 757–763, 1996.

[3] Sharon Barbour. Soldier Andrew Garthwaite's mind-control bionic op, 2012.

[4] JV Basmajian and Carlo J. De Luca. *Muscle Alive. Their Functions Revealed by Electromyography*. Williams & Wilkins, 5 edition, 1985.

[5] Roberto Battiti. Using mutual information for selecting features in supervised neural net learning. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, 5(4):537–50, January 1994.

[6] Anthony J Bell and Terrence J Sejnowski. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. 1159(1994):1129–1159, 1995.

[7] Anthony J Bell and Terrence J Sejnowski. The "independent components" of natural scenes are edge filters. *Vision research*, 37(23):3327–38, December 1997.

[8] Adel Belouchrani, K. Abed-Meraim, J.F. Cardoso, and E. Moulines. Second-order blind separation of temporally correlated sources. In *Proceedings International Conference on Digital Signal Processing*, number 1, pages 346–351, 1993.

[9] Adel Belouchrani, K. Abed-Meraim, J.F. Cardoso, and E. Moulines. A blind source separation technique using second-order statistics. *IEEE Transactions on Signal Processing*, 45(2):434–444, 1997.

[10] Adel Belouchrani and A. Cichocki. SOBI@ICALAB for Signal Processing, 2007.

[11] Julius S. Bendat and Allan G. Piersol. *Random Data*. Wiley, 4th edition, 2010.

[12] Martin Bilodeau, Marco Cincera, A.Bertrand Arsenault, and Denis Gravel. Normality and stationarity of EMG signals of elbow flexor muscles during ramp and step isometric contractions. *Journal of Electromyography and Kinesiology*, 7(2):87–96, 1997.

[13] Christopher M. Bishop. *Pattern recognition and machine learning*. 2006.

[14] Reza Boostani and Mohammad Hassan Moradi. Evaluation of the forearm EMG signal features for the control of a prosthetic hand. *Physiological measurement*, 24(2):309–19, May 2003.

[15] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[16] I. N. Bronštejn, K.A. Semendajev, Gerhard Musiol, and Heiner Mühlig. *Taschenbuch der Mathematik*. 5 edition, 2000.

[17] Gavin Brown. A New Perspective for Information Theoretic Feature Selection. In *12th International Conference on Artificial Intelligence and Statistics*, number 1, pages 49–56, 2009.

[18] Gavin Brown, Adam Pocock, Ming-Jie Zhao, and Mikel Lujan. Conditional Likelihood Maximisation : A Unifying Framework for Information Theoretic Feature Selection. *Journal of Machine Learning Research*, 13:27–66, 2012.

[19] Thomas S. Buchanan, David G. Lloyd, Kurt Manal, and Thor F. Besier. Neuromusculoskeletal modeling: estimation of muscle forces and joint moments and movements from measurements of neural command. *Journal of Applied Biomechanics*, 20(4):367–395, 2004.

[20] a D Chan, K Englehart, B Hudgins, and D F Lovely. Myoelectric signals to augment speech recognition. *Medical & Biological Engineering & Computing*, 39(4):500–4, July 2001.

[21] Adrian D C Chan and Kevin B Englehart. Continuous myoelectric control for powered prostheses using hidden Markov models. *IEEE Transactions on Biomedical Engineering*, 52(1):121–4, January 2005.

[22] Vladimir Cherkassky and Filip Mulier. *Learning from Data*. Wiley, 2 edition, 2007.

[23] Dudley S. Childress. Powered limb prostheses: their clinical significance. *IEEE Transactions on Biomedical Engineering*, (3):200–207, 1973.

[24] Jun-Uk Chu, Inhyuk Moon, Yun-Jung Lee, Shin-Ki Kim, and Mu-Seong Mun. A Supervised Feature-Projection-Based Real-Time EMG Pattern Recognition for Multifunction Myoelectric Hand Control. *IEEE/ASME Transactions on Mechatronics*, 12(3):282–290, June 2007.

[25] Jun-Uk Chu, Inhyuk Moon, and Mu-Seong Mun. A real-time EMG pattern recognition system based on linear-nonlinear feature projection for a multifunction myoelectric hand. *IEEE Transactions on Biomedical Engineering*, 53(11):2232–9, November 2006.

[26] Edward A. Clancy. Electromyogram amplitude estimation with adaptive smoothing window length. *IEEE Transactions on Biomedical Engineering*, 46(6):717–29, June 1999.

[27] Edward A. Clancy. EMG amplitude estimation toolbox: User's guide Alpha version 0.04, 2004.

[28] Edward A. Clancy, Stephane Bouchard, and Denis Rancourt. Estimation and application of EMG amplitude during dynamic contractions. *IEEE Engineering in Medicine and Biology Magazine*, 20(6):47–54, 2001.

[29] Edward A. Clancy and Kristin A. Farry. Adaptive whitening of the electromyogram to improve amplitude estimation. *IEEE Transactions on Biomedical Engineering*, 47(6):709–19, June 2000.

[30] Edward A. Clancy and Neville Hogan. Single site electromyograph amplitude estimation. *IEEE Transactions on Biomedical Engineering*, 41(2), 1994.

[31] Edward A. Clancy and Neville Hogan. Multiple site electromyograph amplitude estimation. *IEEE Transactions on Biomedical Engineering*, 42(2):203–11, February 1995.

[32] Edward A. Clancy and Neville Hogan. Influence of joint angle on the calibration and performance of EMG amplitude estimators. *IEEE Transactions on Biomedical Engineering*, 45(5):664–8, May 1998.

[33] Edward A. Clancy and Neville Hogan. Probability density of the surface electromyogram and its relation to amplitude detectors. *IEEE transactions on biomedical engineering*, 46(6):730–9, June 1999.

[34] Edward A. Clancy, E L Morin, and Rroberto Merletti. Sampling, noise-reduction and amplitude estimation issues in surface electromyography. *Journal of Electromyography and Kinesiology*, 12:1–16, 2002.

[35] Jan Pieter Clarys. Electrology and Localized Electrization. *Journal of Electromyography and Kinesiology*, 4(1), 1994.

[36] T.M. Cover and J.A. Thomas. *Elements of information theory*. Wiley, New York, 2 edition, 2006.

[37] Jeffrey R. Cram. The history of surface electromyography. *Applied Psychophysiology and Biofeedback*, 28(2):81–91, 2003.

[38] Jeffrey R. Cram. *Cram's introduction to surface electromyography*. 2nd edition, 2011.

[39] Jeffrey R. Cram and Jeffrey C. Steger. EMG scanning in the diagnosis of chronic pain. *Biofeedback and Self-regulation*, 8(2):229–41, June 1983.

[40] Scott J. Day and Manuel Hulliger. Experimental Simulation of Cat Electromyogram: Evidence for Algebraic Summation of Motor-Unit Action-Potential Trains. *Journal of Neurophysiology*, 86:2144–2158, 2001.

[41] Carlo J. De Luca. Physiology and mathematics of myoelectric signals. *IEEE Transactions on Biomedical Engineering*, 26(6):313–25, June 1979.

[42] Carlo J. De Luca. The use of surface electromyography in biomechanics. *Journal of Applied Biomechanics*, (July 1993):1–38, 1997.

[43] Catherine Disselhorst-Klug, Jiri Silny, and Günter Rau. Improvement of spatial resolution in surface-EMG: a theoretical and experimental comparison of different spatial filters. *IEEE Transactions on Biomedical Engineering*, 44(7):567–74, July 1997.

[44] K. Englehart, B. Hudgins, and P.a. Parker. Time-frequency based classification of the myoelectric signal: static vs. dynamic contractions. In *Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 317–320. IEEE, 2000.

[45] Kevin Englehart and Bernard Hudgins. A robust, real-time control scheme for multifunction myoelectric control. *IEEE Transactions on Biomedical Engineering*, 50(7):848–54, July 2003.

[46] Kevin Englehart, Bernard Hudgins, P A Parker, and M Stevenson. Classification of the myoelectric signal using time-frequency based representations. *Medical Engineering & Physics*, 21(6-7):431–8, 1999.

[47] Kevin B. Englehart, Bernard S. Hudgins, and Philip A. Parker. A wavelet-based continuous classification scheme for multifunction myoelectric control. *IEEE Transactions on Biomedical Engineering*, 48(3):302–11, March 2001.

[48] Roger M. Enoka. *Neuromechanics of human movement.* 3rd edition, 2002.

[49] Harry B. Evans, Zushan Pan, Philip A. Parker, and Robert N. Scott. Signal Processing for Proportional Myoelectric Control. *IEEE Transactions on Biomedical Engineering*, (2):207–211, 1984.

[50] Dario Farina and Corrado Cescon. Concentric-Ring Electrode Systems for Noninvasive Detection of Single Motor Unit Activity. *IEEE Transactions on Biomedical Engineering*, 48(11):1326–34, November 2001.

[51] Dario Farina, Aleš Holobar, Roberto Merletti, and Roger M. Enoka. Decoding the neural drive to muscles from the surface electromyogram. *Clinical Neurophysiology*, 121(10):1616–1623, 2010.

[52] Dario Farina and Rroberto Merletti. Comparison of algorithms for estimation of EMG variables during voluntary isometric contractions. *Journal of Electromyography and Kinesiology*, 10(5):337–49, October 2000.

[53] Dario Farina, E Schulte, Roberto Merletti, Günter Rau, and Catherine Disselhorst-Klug. Single motor unit analysis from spatially filtered surface electromyogram signals. Part I: spatial selectivity. *Medical & biological engineering & computing*, 41(3):330–7, May 2003.

[54] Todd R Farrell and Richard F Weir. The optimal controller delay for myoelectric prostheses. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 15(1):111–8, March 2007.

[55] Gian Carlo Filligoi and Paolo Mandarini. Some Theoretic Results on a Digital EMG Signal Processor. *IEEE Transactions on Biomedical Engineering*, 31(4):333–341, 1984.

[56] François Fleuret. Fast binary feature selection with conditional mutual information. *The Journal of Machine Learning Research*, 5:1531–1555, 2004.

[57] Emma Fortune and Madeleine M Lowery. Effect of extracellular potassium accumulation on muscle fiber conduction velocity: a simulation study. *Annals of Biomedical Engineering*, 37(10):2105–17, October 2009.

[58] Anders Fougner, Erik Scheme, Adrian D C Chan, Kevin Englehart, and Oyvind Stavdahl. Resolving the limb position effect in myoelectric pattern recognition. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 19(6):644–51, December 2011.

[59] Yoav Freund and Robert E Schapire. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1):119–139, August 1997.

[60] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*. 2 edition, 2001.

[61] GA García, Kazuya Maekawa, and Kenzo Akazawa. Decomposition of synthetic multi-channel surface-electromyogram using independent component analysis. *Independent Component Analysis, LNCS*, pages 985–992, 2004.

[62] MR Garey and DS Johnson. "Strong" NP-Completeness Results: Motivation, Examples, and Implications. *Journal of the ACM*, 25(3):499–508, 1978.

[63] H. Gävert, J. Hurri, J. Särelä, and Aapo Hyvärinen. The FastICA for package for MATLAB, 2005.

[64] Daniel Graupe and William K. Cline. Functional separation of EMG signals via ARMA identification methods for prosthesis control purposes. *Systems, Man and Cybernetics, IEEE Transactions on*, 2:252–259, 1975.

[65] Moritz Grosse-Wentrup and Martin Buss. Multiclass common spatial patterns and information theoretic feature extraction. *IEEE Transactions on Biomedical Engineering*, 55(8):1991–2000, August 2008.

[66] Isabelle Guyon. *Feature extraction: foundations and applications*. Springer Verlag, 2006.

[67] Janne Hahne, Bernhard Graimann, and Klaus-Robert Müller. Spatial Filtering for Robust Myoelectric Control. *IEEE Transactions on Biomedical Engineering*, 59(5):1–8, 2012.

[68] L Hargrove, K Englehart, and B Hudgins. A training strategy to reduce classification degradation due to electrode displacements in pattern recognition based myoelectric control. *Biomedical Signal Processing and Control*, 3(2):175–180, April 2008.

[69] L. Hargrove, E. Scheme, Kevin B. Englehart, and Bernard Hudgins. Principal Components Analysis Preprocessing to Reduce Controller Delays in Pattern Recognition Based Myoelectric Control. In *Proceedings of the 29th International Conference of the IEEE EMBS*, pages 6511–6514, 2007.

[70] H J Hermens, B Freriks, C Disselhorst-Klug, and G Rau. Development of recommendations for SEMG sensors and sensor placement procedures. *Journal of Electromyography and Kinesiology*, 10(5):361–74, October 2000.

[71] David Hofmann and J. Michael Herrmann. Spatial Feature Extraction for Classification of Nonstationary Myoelectric Signals. In *11th International Conference on Machine Learning and Applications*, volume 2, pages 416–421. IEEE, December 2012.

[72] Neville Hogan and Robert W Mann. Myoelectric signal processing: Optimal estimation applied to electromyography-Part I: Derivation of the optimal myoprocessor. *IEEE Transactions on Biomedical Engineering*, 27(7):382–395, 1980.

[73] Neville Hogan and Robert W Mann. Myoelectric signal processing: Optimal estimation applied to electromyography-part ii: experimental demonstration of optimal myoprocessor performance. *IEEE Transactions on Biomedical Engineering*, (7):396–410, 1980.

[74] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441, 1933.

[75] He Huang, Ping Zhou, Guanglin Li, and Todd Kuiken. Spatial filtering improves EMG classification accuracy following targeted muscle reinnervation. *Annals of Biomedical Engineering*, 37(9):1849–57, September 2009.

[76] Yonghong Huang, Kevin B Englehart, Bernard Hudgins, and Adrian D C Chan. A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses. *IEEE Transactions on Biomedical Engineering*, 52(11):1801–11, November 2005.

[77] Bernard S Hudgins, Philip A. Parker, and Robert N. Scott. A new strategy for multifunction myoelectric control. *IEEE Transactions on Biomedical Engineering*, 40(1):82–94, 1993.

[78] I. W. Hunter, R. E. Kearney, and L. A. Jones. Estimation of the conduction velocity of muscle action potentials using phase and impulse response function techniques. *Medical & Biological Engineering & Computing*, 25(2):121–126, March 1987.

[79] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent component analysis*. Wiley, 2001.

[80] Verne T. Inman, H. J. Ralston, J. B. de C. M. Saunders, Bertram Feinstein, and Elwood W. Writght. Relation of human electromyogram to muscular tension. *Electroencephalography and Clinical Neurophysiology*, 4(2):187–194, 1952.

[81] Abhishek Jaiantilal. Random Forests implementation for MATLAB.

[82] Ben H Jansen, Vonda H Miller, Demetrios C Mavrofrides, and Caroline W Stegink Jansen. Multidimensional EMG-based assessment of walking dynamics. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(3):294–300, September 2003.

[83] Ning Jiang, Strahinja Dosen, Klaus-Robert Müller, and Dario Farina. Myoelectric control of artificial limbs: is there the need for a change of focus? *IEEE Signal Processing Magazine*, 152:12–15, 2012.

[84] I. T. Jolliffe. *Principal component analysis*. 2 edition, 2002.

[85] Jari Kaipo and Erkki Somersalo. *Statistical and Computational Inverse Problems*. Springer, 2004.

[86] ER Kandel, JH Schwartz, and TM Jessell. *Principles of neural science*. 2000.

[87] Kevin G Keenan, Dario Farina, Katrina S Maluf, Roberto Merletti, and Roger M Enoka. Influence of amplitude cancellation on the simulated surface electromyogram. *Journal of Applied Physiology*, 98(1):120–31, January 2005.

[88] Kevin G. Keenan and Francisco J. Valero-Cuevas. Epoch length to accurately estimate the amplitude of interference EMG is likely the result of unavoidable amplitude cancellation. *Biomedical Signal Processing and Control*, 3(2):154–162, 2008.

[89] Cynthia Kendell, Edward D Lemaire, Yves Losier, Adam Wilson, Adrian Chan, and Bernie Hudgins. A novel approach to surface electromyography: an exploratory study of electrode-pair selection based on signal characteristics. *Journal of neuroengineering and rehabilitation*, 9:24, January 2012.

[90] J. Kennedy and R. Eberhart. Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks*, 4:1942–1948, 1995.

[91] James Kennedy and RC Eberhart. A discrete binary version of the particle swarm algorithm. In *Systems, Man, and Cybernetics*, pages 4–8, 1997.

[92] Mahdi Khezri and Mehran Jahed. An exploratory study to design a novel hand movement identification system. *Computers in biology and medicine*, 39(5):433–442, 2009.

[93] Rf Kleissen, Jh Buurke, J Harlaar, and G Zilvold. Electromyography in the biomechanical analysis of human movement and its clinical application. *Gait & posture*, 8(2):143–158, October 1998.

[94] R. Benjamin Knapp and Hugh S. Lusted. A bioelectric controller for computer music applications. *Computer Music Journal*, 14(1):42–47, 1990.

[95] Z. J. Koles. The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG. *Electroencephalography and Clinical Neurophysiology*, 79(6):440–7, December 1991.

[96] Samuel Kotz, Tomasz J. Kozubowski, and Podgorski Krzysztof. *The Laplace Distribution and Generalizations*. Birkh{ä}user, Boston, 2001.

[97] John G. Kreifeldt. Signal versus noise characteristics of filtered EMG used as a control source. *IEEE Transactions on Biomedical Engineering*, BME-18(1):16–22, 1971.

[98] John G. Kreifeldt and Sumner Yao. A Signal-to-Noise Investigation of Nonlinear Electromyographic Processors. *IEEE Transactions on Biomedical Engineering*, 21(4):298–308, 1974.

[99] Todd A. Kuiken, Madeleine M. Lowery, and Nikolay S. Stoykov. The effect of subcutaneous fat on myoelectric signal amplitude and cross-talk. *Prosthetics and Orthotics International*, 27(1):48–54, April 2003.

[100] Eugene Kwatny, Donald H. Thomas, and Harry G. Kwatny. An application of signal processing techniques to the study of myoelectric signals. *IEEE Transactions on Biomedical Engineering*, Vol. BME-1(4):303–313, 1970.

[101] J. H. Lawrence and Carlo J. De Luca. Myoelectric signal versus force relationship in different human muscles. *Journal of applied physiology: respiratory, environmental and exercise physiology*, 54(6):1653–9, June 1983.

[102] E.L. Lehmann and George Casella. *Theory of point estimation*. Springer, 2 edition, 1998.

[103] Thomas Lorrain, Ning Jiang, and Dario Farina. Influence of the training set on the accuracy of surface EMG classification in dynamic contractions for the control of multifunction prostheses. *Journal of NeuroEngineering and Rehabilitation*, 8(1):25, 2011.

[104] Marie-Francoise Lucas, Adrien Gaufriau, Sylvain Pascual, Christian Doncarli, and Dario Farina. Multi-channel surface EMG classification using support vector machines and signal-based wavelet optimization. *Biomedical Signal Processing and Control*, 3(2):169–174, 2008.

[105] David J.C. MacKay. *Information theory, inference, and learning algorithms*. Cambridge University Press, Cambridge UK ;;New York, 2003.

[106] Roberto Merletti and Dario Farina. Surface EMG processing: Introduction to the special issue. *Biomedical Signal Processing and Control*, 3(2):115–117, April 2008.

[107] Roberto Merletti, M Knaflitz, and Carlo J. De Luca. Electrically evoked myoelectric signals. *Critical reviews in biomedical engineering*, 19(4):293–340, January 1992.

[108] Roberto Merletti and Philip A. Parker. *Electromyography*. Wiley, 1 edition, 2004.

[109] H. S. Milner-Brown and R. B. Stein. The relation between the surface electromyogram and muscular force. *The Journal of Physiology*, 246:549–569, 1975.

[110] Tom M. Mitchell. *The discipline of machine learning*. 2006.

[111] Ganesh Naik and Dinesh Kumar. Hybrid Feature Selection for Myoelectric Signal Classification Using MICA. *Journal of Electrical Engineering*, 61(2):93–99, January 2010.

[112] Ganesh R Naik, Dinesh K Kumar, and Hans Weghorn. ICA based identification of sources in sEMG. *2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information*, pages 619–624, 2007.

[113] Hideo Nakamura, Masaki Yoshida, Manabu Kotani, Kenzo Akazawa, and Toshio Moritani. The application of independent component analysis to the multi-channel surface electromyographic signals for separation of motor unit action potential trains: part I-measuring techniques. *Journal of Electromyography and Kinesiology*, 14(4):423–32, August 2004.

[114] Kianoush Nazarpour, Ahmad R. Sharafat, and S. Mohammad P. Firoozabadi. Application of higher order statistics to surface electromyogram signal classification. *IEEE Transactions on Biomedical Engineering*, 54(10):1762–1769, October 2007.

[115] Francesco Negro, Ales Holobar, and Dario Farina. Fluctuations in isometric muscle force can be described by one linear projection of low-frequency components of motor unit discharge rates. *The Journal of Physiology*, 587(Pt 24):5925–38, December 2009.

[116] Philip Nelson. *Biological Physics: Energy, Information, Life*. 2008.

[117] Andrew Y. Ng. *On feature selection: learning with exponentially many irrevelant features as training examples*. Master thesis, Massachusetts Institute of Technology, 1998.

[118] Anthony O'Hagan. *Kendall's Advanced Theory of Statistics 2B: Bayesian Inference*. 1994.

[119] Mohammadreza Asghari Oskoei and Huosheng Hu. Support vector machine-based classification scheme for myoelectric control applied to upper limb. *IEEE transactions on bio-medical engineering*, 55(8):1956–65, August 2008.

[120] Mohammadreza Asghari Oskoei and Huosheng Hu. Adaptive myoelectric human-machine interface for video games. *2009 International Conference on Mechatronics and Automation*, pages 1015–1020, August 2009.

[121] Nils Ostlund, J Yu, K Roeleveld, and J S Karlsson. Adaptive spatial filtering of multichannel surface electromyogram signals. *Medical & biological engineering & computing*, 42(6):825–31, December 2004.

[122] Nils Ostlund, Jun Yu, and J Stefan Karlsson. Adaptive spatio-temporal filtering of multichannel surface EMG signals. *Medical & biological engineering & computing*, 44(3):209–15, March 2006.

[123] Euljoon Park and Sanford G. Meek. Adaptive filtering of the electromyographic signal for prosthetic control and force estimation. *IEEE Transactions on Biomedical Engineering*, 42(10):1048–1052, 1995.

[124] PA Parker, JA Stuller, and RN Scott. Signal processing for the multistate myo-electric channel. *Proceedings of the IEEE*, 65(5), 1977.

[125] Philip A. Parker, Kevin B. Englehart, and Bernard S. Hudgins. Myoelectric signal processing for control of powered limb prostheses. *Journal of Electromyography and Kinesiology*, 16(6):541–8, December 2006.

[126] K Pearson. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine*, 2:559–572, 1901.

[127] Hanchuan Peng, Fuhui Long, and Chris Ding. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE transactions on pattern analysis and machine intelligence*, 27(8):1226–38, August 2005.

[128] Riccardo Poli, James Kennedy, and Tim Blackwell. Particle swarm optimization. *Swarm Intelligence*, 1(1):33–57, August 2007.

[129] J R Potvin and S H M Brown. Less is more: high pass filtering, to remove up to 99% of the surface EMG signal power, improves EMG-based biceps brachii muscle force estimates. *Journal of Electromyography and Kinesiology*, 14(3):389–99, June 2004.

[130] Punit Prakash, Christian A. Salini, John A. Tranquilli, Donald R. Brown, and Edward A. Clancy. Adaptive whitening in electromyogram amplitude estimation for epoch-based applications. *IEEE Transactions on Biomedical Engineering*, 52(2):331–334, 2005.

[131] P Pudil, J Novovicova, and J Kittler. Floating search methods in feature selection. *Pattern recognition letters*, 15(November):1119–1125, 1994.

[132] Christopher L. Pulliam, Joris M. Lambrecht, and Robert F. Kirsch. Electromyogram-based neural network control of transhumeral prostheses. *The Journal of Rehabilitation Research and Development*, 48(6):739, 2011.

[133] M B I Raez, M S Hussain, and F Mohd-Yasin. Techniques of EMG signal analysis: detection, processing, classification and applications. *Biological procedures online*, 8(1):11–35, January 2006.

[134] Narender P Reddy and Vineet Gupta. Toward direct biocontrol using surface EMG signals: control of finger and wrist joint models. *Medical Engineering & Physics*, 29(3):398–403, April 2007.

[135] Francesco Redi. Esperienze intorno a diverse cose naturali e particolarmente a quelle che ci sono portate dalle Indie. 1671.

[136] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958.

[137] Nina Rudigkeit. *Robuste myoelektrische Signalverarbeitung*. PhD thesis, 2012.

[138] Terence D Sanger. Bayesian filtering of myoelectric signals. *Journal of neurophysiology*, 97(2):1839–45, February 2007.

[139] T Scott Saponas, Desney S. Tan, Dan Morris, and Ravin Balakrishnan. Demonstrating the feasibility of using forearm electromyography for muscle-computer interfaces. *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*, page 515, 2008.

[140] Massimo Sartori, David G Lloyd, Monica Reggiani, and Enrico Pagello. Fast operation of anatomical and stiff tendon neuromuscular models in EMG-driven modeling. *2010 IEEE International Conference on Robotics and Automation*, pages 2228–2234, May 2010.

[141] Marc H. Schieber. Muscular production of individuated finger movements: the roles of extrinsic finger muscles. *The Journal of Neuroscience*, 15:284–297, 1995.

[142] C E Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27:379–423,623–656, 1948.

[143] Yves St-Amant, Denis Rancourt, and Edward A. Clancy. Influence of smoothing window length on electromyogram amplitude estimates. *IEEE Transactions on Biomedical Engineering*, 45(6):795–800, June 1998.

[144] D Staudenmann, I Kingma, A Daffertshofer, D F Stegeman, and Jaap H van Dieën. Heterogeneity of muscle activation in relation to force direction: a multi-channel surface electromyography study on the triceps surae muscle. *Journal of Electromyography and Kinesiology*, 19(5):882–95, October 2009.

[145] Didier Staudenmann, Idsart Kingma, Andreas Daffertshofer, and Dick F Stegeman. Improving EMG-Based Muscle Force Estimation by Using a High-Density EMG Grid and Principal Component Analysis. *IEEE Transactions on Biomedical Engineering*, 53(4):712–719, 2006.

[146] Didier Staudenmann, Idsart Kingma, Dick F Stegeman, and Jaap H van Dieën. Towards optimal multi-channel EMG electrode configurations in muscle force estimation: a high density EMG study. *Journal of Electromyography and Kinesiology*, 15:1–11, 2005.

[147] Didier Staudenmann, Karin Roeleveld, Dick F Stegeman, and Jaap H van Dieën. Methodological aspects of SEMG recordings for force estimation–a tutorial and review. *Journal of Electromyography and Kinesiology*, 20(3):375–87, June 2010.

[148] M Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):111–147, 1974.

[149] Abdulhamit Subasi and M Kemal Kiymik. Muscle fatigue detection in EMG using time-frequency methods, ICA and neural networks. *Journal of medical systems*, 34(4):777–85, August 2010.

[150] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press Cambridge, USA, 2nd edition, 2012.

[151] Michele Tinazzi, Ina Juergenson, Giovanna Squintani, Gaetano Vattemi, Stefania Montemezzi, Daniela Censi, Paolo Barone, Tommaso Bovi, and Alfonso Fasano. Pisa syndrome in Parkinson's disease: an electrophysiological and imaging study. *Journal of Neurology*, 260(8):2138–48, August 2013.

[152] L. Tong, V.C. Soon, Y. F. Huang, and R. Liu. AMUSE: a new blind identification algorithm. *Proceedings of International Symposium on Circuits and Systems*, pages 1784–1787, 1990.

[153] Francisco J. Valero-Cuevas, F. E. Zajac, and C. G. Burgar. Large index-fingertip forces are produced by subject-independent patterns of muscle excitation. *Journal of Biomechanics*, 31(8):693–703, August 1998.

[154] Vladimir N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 2 edition, 1999.

[155] GB Whatmore and DR Kohli. *The Physiopathology and Treatment of Functional Disorders: Including Anxiety States and Depression and the Role of Biofeedback Training*. 1974.

[156] M. E. Wootten, M. P. Kadaba, and G. V. B. Cochran. Dynamic electromyography. I. Numerical representation using principal component analysis. *Journal of Orthopaedic Research*, 8(2):247–58, March 1990.

[157] Chau H. Wu. Electric fish and the discovery of animal electricity. *American Scientist*, 72(6):598–607, 1984.

[158] Howard Hua Yang and John Moody. Feature selection based on joint mutual information. In *Proceedings of International ICSC Symposium on Advances in Intelligent Data Analysis*, pages 22–25, 1999.

[159] Mahyar Zardoshti-Kermani, Bruce C. Wheeler, Kambiz Badie, and Reza M. Hashemi. EMG feature evaluation for movement control of upper extremity prostheses. *IEEE Transactions on Rehabilitation Engineering*, 3(4):324–333, 1995.

[160] Elana M. Zion Golumbic, Nai Ding, Stephan Bickel, Peter Lakatos, Catherine A. Schevon, Guy M. McKhann, Robert R. Goodman, Ronald Emerson, Ashesh D. Mehta, Jonathan Z. Simon, David Poeppel, and Charles E. Schroeder. Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a "Cocktail Party". *Neuron*, 77(5):980–991, 2013.

**OvO** one-vs-one

**OvR** one-vs-rest

**ICA** independent component analysis

**CSP** common spatial patterns

**LDA** linear discriminant analysis

**EM** expectation maximization

**PCA** principal component analysis

**SOBI** second order blind identification

**CSP** common spatial patterns

**DPSO** discrete particle swarm optimization

**PSO** particle swarm optimization

**SFS** sequential forward selection

**SFFS** sequential forward floating selection

**SNR** signal-to-noise ratio

**MES** myoelectric signal

**EMG** electromyogram

**EEG** electroencephalogram

**sEMG** surface electromyogram

**RMS** root mean square

**MAV** mean absolute value

**RMSE** root mean square error

**MU** motor unit

**AP** action potential

**ARMA** auto-regressive moving average

**MUAP** motor unit action potential

**MUAPT** motor unit action potential train

**BFP** Bayes-Fokker-Planck

**jmi** joint mutual information

**mifs** mutual information feature selection

**cmim** conditional mutual information maximization

**mrmr** maximum relevance minimum redundancy