Czech technical univeristy in Prague

Faculty of electrical engineering
Department of cybernetics

Deep Brain Recordings in Parkinson's Disease: Processing, Analysis and Fusion with Anatomical Models

Doctoral Thesis

Ing. Eduard Bakštein

Prague, October 2016

PhD. Study Programme: P2612 - Electrotechnics and Informatics
Branch of Study: 3902V035 - Artificial Intelligence and Biocybernetics

Supervisor: **Ing. Daniel Novák, Ph.D.**
Supervisor-Specialist: **Prof. RNDr. Olga Štěpánková, CSc.**

Abstract

This thesis presents several novel techniques and tools for automatic classification and analysis of highly detailed invasive recordings of the brain activity in patients with Parkinson's disease (PD). By utilizing machine learning concepts, we approach three of the principal questions, central to modern treatment and understanding of the PD:

i) What information about patient's state can be derived from recorded brain activity? By identifying patterns characteristic for tremor onset in signals recorded through deep brain stimulation electrodes, we show that an adaptive system, modifying treatment parameters to match current state of its bearer, is feasible.

ii) How to obtain trustworthy answers to scientific questions from noisy microelectrode activity recordings? We show that undesirable noise is highly prevalent in intraoperative microelectrode recordings and provide the sigInspect: a GUI tool for annotation of microelectrode signals. The tool includes a set of well-performing classifiers for automatic artifact identification, validated on an extensive multi-center database of manually labeled data.

iii) Where exactly in the target nucleus were the signals recorded? This question is vital for appropriate stimulation electrode placement as well as for better understanding of possible side effects. We propose a novel probabilistic model for fitting a 3D anatomical atlas of the subthalamic nucleus based solely on the recorded electrophysiological activity and show that such approach may lead to more accurate localization of recording sites during and after the surgery.

Keywords: microelectrode recordings, machine learning, anatomical model fitting, deep brain stimulation, Parkinson's disease

Declaration

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Prague, Czech Republic
October 2016

Eduard Bakštein

Acknowledgment

I would like to express my many thanks and sincere gratitude to the following people, who helped me during the work on this thesis:

- Daniel Novák and Olga Štěpánková – my supervisors – for their guidance and constant support throughout the work on this thesis as well as during my entire studies.
- Tomáš Sieger, Jiří Wild, Jiří Anýž, Pavel Vostatek and Jakub Schneider – my colleagues and fellow students – especially for many fruitful discussions that helped to form my research (in which most of them participated in some way), as well as for a lot of inspiration and fun.
- Robert Jech for providing the neuroscientific and clinical insight and for steering my research towards problems relevant to the medical research community.
- The team from the Dept. of Neurology and Center of Clinical Neuroscience, Charles University in Prague and the Dept. of Stereotactic and Radiation Neurosurgery, Na Homolce Hospital, Prague: Filip Růžička, Daniela Šťastná, Markéta Fialová and Dušan Urgošík, for providing the data and helping me understand details of the surgical procedures, treatment procedures and the data itself.
- Kevin Warwick, Jon Burgess and the team from the University of Reading, for collaboration and a lot of help with the tremor detection research.
- Marie – my dear wife – for her immense support, patience and everything else.

Contents

List of Figures

List of Tables

List of Abbreviations

PD	Parkinson's disease
DBS	Deep brain stimulation
LFP	Local field potentials
MER	Micro-electrode recordings
μEEG	Micro electro-encephalogram (i.e. MER)
STN	Subthalamic nucleus
EMG	Electromyography
FFT	Fast Fourier transform
PSD	Power spectral density
PDF	Probability density function
MLE	Maximum likelihood estimate
HMM	Hidden Markov models

Chapter 1

Introduction

Over the past decades, an increasing number of technological improvements in medicine provide relief to patients with previously untreatable diagnoses. This trend is made possible by the constant progress in microelectronics, as well as by the rapid development of machine learning, signal processing, and data analysis methods. This thesis presents several contributions to one candidate of this progress: the deep brain stimulation (DBS) for Parkinson's disease. For almost two decades, the DBS helps patients with late-stage Parkinson's disease in cases in which the medication fails to control their symptoms sufficiently. In many severe cases, the DBS helps the patients to re-establish balance to an unforeseen extent and increases their quality of life considerably.

To achieve a good clinical outcome with low side-effects, the stimulation electrode has to be placed to a particular area in the basal ganglia — a structure deep in the brain. The prevailing technique to attain the necessary accuracy is based on recording of neuronal activity in the neighborhood of the expected target location. Primarily, these recordings are used by a trained expert to identify signal properties characteristic for the target area during the surgery. Additionally, the procedure provides a unique possibility to record neuronal activity from deep structures in the human brain in a very high detail — up to the level of activity of individual neurons. This allows the research community to investigate the function of the human brain, as well as mechanisms of the DBS and the disease itself.

This thesis attempts to aid improving accuracy and efficacy in DBS and neuroscientific research by applying machine learning techniques to several steps of the process: the Chapter 3 suggests the possibility of automatic detection of Parkinsonian tremor onset from neuronal activity, recorded directly from the stimulation electrodes. This way, the implanted stimulator could be switched on automatically in moments when the symptoms emerge, which might lead to battery conservation with the benefit of a prolonged period before device replacement. Another possible benefit would be a more delicate treatment of the brain tissue due to reduced stimulation time, possibly leading to increased period of DBS efficacy.

As the primary purpose of the DBS is treatment, the signals recorded during the procedure contain a large amount of external artifacts, introduced by electromagnetic inference or mechanical vibrations. The Chapter 4 provides a model for automatic artifact detection, which is evaluated on an extensive database of annotated microelectrode recordings. Furthermore an interactive tool for signal inspection and automatic annotation is presented.

During the DBS surgery and electrode implantation, creating a probabilistic model of neural activity parameters along electrode trajectories (Chapter 5), may bring objective and accurate identification of the target area. In the commonly used multi-electrode trajectories, this approach may even be extended to fitting an anatomical model of the target nucleus to the microelectrode recordings directly (Chapter 6). Accurate localization of electrode position within the target nucleus is vital for achieving good clinical outcome and providing a three-dimensional model may lead not only to improved targeting accuracy or decreased surgical time but also contribute to overall understanding of the DBS technique and the Parkinson's disease itself.

1.1 Goals of the thesis

The main aims of this thesis include:

- To investigate the possibility to detect Parkinson's disease tremor onset solely from neuronal activity recorded using the implanted stimulation electrode. This effort may lead to an adaptive system, modifying stimulation parameters to match the current state of its bearer.
- To develop a system for automatic artifact identification in the microelectrode recordings, focusing on externally induced noise. If present, the artifacts may negatively affect the subsequent analysis and a system for their automatic detection may bring ease to data preprocessing and increased sensitivity to answering scientific questions.
- To develop an automatic model for classification of brain nuclei along electrode trajectories in the deep brain stimulation surgery, based on microelectrode recordings, captured during electrophysiological exploration. Automatic classification may increase accuracy and speed of target localization and contribute to increased therapeutic outcome.
- To investigate the possibility to fit a three-dimensional anatomical model of the subthalamic nucleus directly to the microelectrode data, recorded along a set of parallel exploration microelectrodes. Finding the most likely position of recording sites within a 3D model of the target nucleus may increase accuracy, efficacy, and understanding of the DBS procedure.

Chapter 2

Background

Chapter summary

The main aim of this chapter is to provide the reader with a very brief introduction to the Deep Brain Stimulation (DBS) — a modern treatment method for Parkinson's Disease (PD) — which relies on precise electrode placement into the deep brain structures. During the surgery, as well as during the subsequent therapy, the DBS provides a unique insight into the function of the human central nervous system by enabling direct high-resolution recordings of the neural activity. This chapter provides summary of the procedure, as well as of the two main types of signals that can be obtained:

1. The microelectrode recordings (MER), recorded during the surgical procedure using fine microelectrodes, capable of recording single neuron's activity.

2. The local field potentials (LFP), recorded typically post-operatively using large stimulation electrodes, capturing activity of larger neuronal populations, suitable for long-term monitoring.

2.1 Parkinson's disease

The PD is a common neurodegenerative chronic disorder, affecting most commonly elderly people, which is estimated to affect about 6.3 million people worldwide (*European Parkinson's Disease Association* 2016). The incidence of PD rises substantially with age (mean age at diagnosis around 70 years, only about ten percent cases diagnosed before the age of 50) and the incidence appears to be slightly higher in men than in women (Van Den Eeden 2003).

The major symptoms are motorical and include rigidity (muscle stiffness), bradykinesia (slowness of movement), tremor (shaking movement of body extremities) and postural instability (useful overview of PD symptoms can be found in (Jankovic 2008)). The cause is impaired function of the Basal ganglia — a neurological structure at the base of the forebrain, responsible mainly for control of voluntary movements. Other processes in which the basal ganglia play an important role include control of eye movements, procedural learning and cognitive and emotional processes.

In PD, dopamine producing cells in the Substantia nigra (SNr) become necrotic, resulting in reduced production of this neurotransmitter. The lack of dopamine then introduces imbalance into the complicated system of excitatory and inhibitory neuronal pathways between the nuclei, which leads to impaired function of the basal ganglia (Obeso et al. 2011). Despite the high number of factors that are suspected to contribute to PD development (Olanow et al. 1999), the actual cause of this process is still unknown, as well as there is no recognized cure for PD.

2.1.1 PD treatment

As no cure for the degenerative process exists to date, the therapeutic methods currently used aim at alleviation of the PD symptoms. Most common therapy involves regular application of levodopa (or L-dopa), a dopamine precursor that is transformed to dopamine in the body. Despite the high proportion of PD patients benefiting from levodopa, several reasons exist that make levodopa problematic for long-term users. As the PD is a chronic disorder, many of the sufferers live with the disease for many years. Over time, the beneficial effects of levodopa may diminish and increased doses are required to achieve sufficient level of symptom suppression. The long-term therapy together with increasing doses often result in rising amount of negative side-effects, such as dyskinesias, toxicity or so-called "on-off switching" when the beneficial effect of drug therapy alternates with episodes of severe PD symptoms. Problems with long-term levodopa therapy are very frequent and may affect as much as 75% of chronic PD patients after 6 years of therapy (Fahn 2006). The efficacy of long-term therapy may be improved by using combination of medications, including dopamine agonists[1] and careful dosage but the

[1] chemical agents that stimulate the dopamine receptors directly

quality of life declines over time in a vast majority of PD patients (Fahn 2006; Marsden 1994).

In these long-term PD cases where levodopa therapy does not provide appropriate treatment anymore, electrical stimulation of the basal ganglia structures — the DBS — can be used (Limousin et al. 1998; Benabid 2003). Although a connection exists between pre-operative response to medication and outcome of the STN DBS stimulation (Pahwa et al. 2006), the beneficial therapeutic effect in great number of PD patients lead to broad use of this technique worldwide.

2.2 Deep brain stimulation

Since the beginnings of human DBS treatment in the 1990's (Alesch et al. 1995; Limousin et al. 1998), the DBS has proved to provide significant improvement for advanced PD patients (Krack et al. 2003) and established as an FDA-approved standard PD treatment, applied in hundreds of neurological centers worldwide (Benabid 2003; Bakay et al. 2011; Abosch et al. 2013). Apart from parkinsonism, DBS is commonly used also for treatment of clinical depressions, dystonia, essential tremor or obsessive-compulsive disorder (Farris et al. 2011). In this report, however, we will refer to the DBS technique exclusively with respect to the PD.

The DBS is based on application of electrical pulses through electrodes surgically implanted into patient's brain. During the surgical procedure, target nuclei (compact brain structures) are identified using Magnetic Resonance Imaging (MRI) and electrophysiological exploration. Their position is determined according to stereotactic frame, attached to patient's head and the stimulation electrodes are then carefully implanted to the identified target. Once the transient effects, caused by oedema and microlesion in the target subside, stimulation electrodes are connected to the stimulator device, implanted in the chest cavity. All leads are internalized and all further adjustments to the stimulatory parameters are done remotely through the skin. The stimulator device itself closely resembles heart pacemaker and uses similar titanium casing and terminals. Schematic overview of a DBS system is shown in Figure 2.1.

The choice of target structure depends on disease type and observed symptoms. In case of the PD, the most commonly used target structures are within the basal ganglia and include the Subthalamic Nucleus (STN) and the Globus Pallidus (GPi), the former being more prevalent (Gross et al. 2006). Common stimulation signals are rectangular pulses at frequencies in the range around 100-200 Hz (Benabid et al. 1996). Actual stimulation parameters such as frequency, pulse width and amplitude are fine-tuned for each patient in order to gain maximum therapeutic outcome (Benabid 2003). The stimulation parameters are set by the caregiving physician in a series of experiments, using a remote programming device. This way, ambulatory adjustments may be done at any time during the treatment with no need of surgical procedures

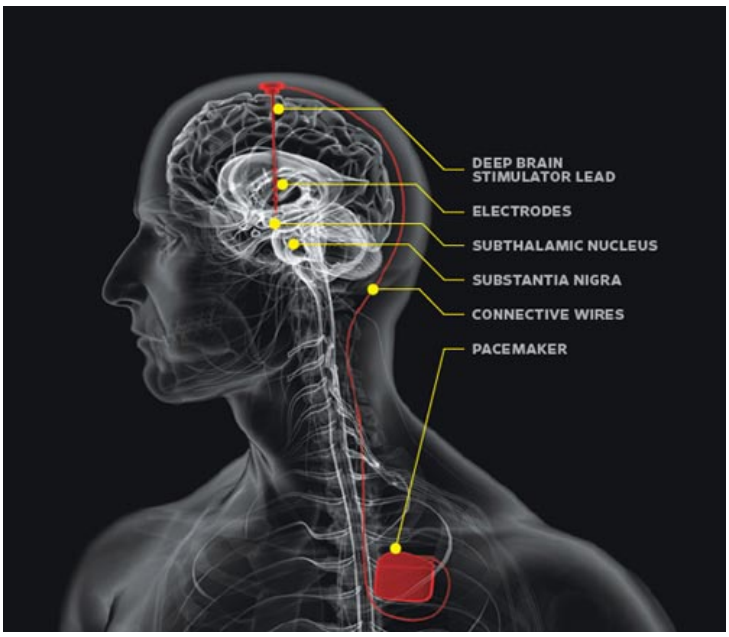

Figure 2.1: Overview of a DBS system, taken from (*Human Brain Stimulation and Electrophysiology Lab, Department of Psychology, University of Texas at Austin* 2012)

in order to match stimulatory parameters to current patient's state.

2.2.1 Optimization of stimulation parameters

The standard prcedure relies on subjective evaluation of symptom severity using the unified Parkinson's disease rating scale (Goetz et al. 2007). In this procedure, the neurologist strives to achieve maximal reduction in symptom severity by adjustments to stimulation parameters and medication. Over the past years, efforts have been made to gain more understanding of optimal stimulation parameter values and their dependence on clinical record or objectively using accelerometers (Pulliam et al. 2015).

As the stimulation electrode itself can form a source of valuable signal, some researchers aimed at real-time adaptation of stimulation parameters based on current state of the patient, pursuing switch to a *closed-loop* system with adaptive parameters (McIntyre et al. 2015). One such approach — detection of tremor onset — can be found in Chapter 3 of this thesis. Thanks to the fact that the local field potential signals (see description below) are relatively stable over the years of therapy (Abosch et al. 2012), such approach might bring benefit to PD patients in the form of reduced therapeutical cost (thanks to lower frequency of device replacements due to depleted battery), as well as improved therapeutical outcome (thanks to parameters optimized to current clinical state of the patient.)

2.2.2 Targetting in the DBS surgery

To achieve a good outcome of the DBS therapy, accurate localization of the target structure is necessary. This is a very complicated task not only due to the position of the basal ganglia, surrounded by a thick layer of forebrain, but especially due to small dimensions of the target nuclei. In case of STN, the dimensions are around and below 10 mm (Daniluk et al. 2010), which makes accurate-enough localization of its boundaries difficult and problematic. A study using standard Magnetic Resonance Imaging (MRI) targetting reported best achieved implantation accuracy about 2 mm (Rezai et al. 2006). The inacurracy in purely imaging-based localization is caused by targeting errors related to the inherent resolution limitations of neuroimaging (standard voxel size is around $0.5 \times 0.5 \times 2$ mm in T2-weighted 1.5T MRI) as well as anatomic shifts during surgery (Halpern et al. 2008; Nimsky et al. 2000). Therefore, once standard and broadly used methods such as matching of the MRI images to brain atlases (e.g. in Castro et al. 2006) are surpassed by their combination with electrophysiological exploration (Simon et al. 2005; Tarsy et al. 2008). Additionaly to planning surgery systems, developed by manufacturers of surgical instrumentation, sophisticated systems to aid in DBS trajectory planning, visualization and electrode positioning have also been developed in the research community (Guo et al. 2005; Miocinovic et al. 2007; Essert et al. 2012; Horn et al. 2014). While the resolution, contrast and overall accuracy of pre-operative imaging may be improved in the future by the use of ultra-high field 7T MRI scanning (Duchin et al. 2012), the use of MER exploration will probably remain part of the DBS stereotactic procedure in the upcoming years.

While a small proportion of centers implant the DBS based solely on anatomy, the vast majority include some form of physiological mapping to define the optimal site, including microelectrode recording (MER), microstimulation, and/or macrostimulation testing (Gross et al. 2006). The use of electrophysiological exploration to identify precise position of the target has been accepted as a recommended setup already for a long time (Rezai et al. 2006). A recent worldwide study found that about 83% of the surveyed centers used MER for position refinement (Abosch et al. 2013).

In MER targetting, the mapping process commonly involves identification of the entry and exit points of the STN across the MER electrode trajectories. Additionally, localization of the sensorimotor area whithin the STN with distinct beta-band oscillations may be performed to utilize the reported influence of exact positioning of the stimulation contact within the STN (Zonenshayn et al. 2004; Zaidel et al. 2010; Verhagen et al. 2015). To obtain a complete spatial information about target boundaries, multiple electrode trajectories are often used. Each MER pass carries with it the risk of hemorrhagic complications (Gorgulho et al. 2005; Rezai et al. 2006; Hariz et al. 1999), and increased surgical time. Therefore, more sophisticated localization methods may improve the speed, accuracy, and safety of DBS implantation. Many

researchers suggested that automated methods of MER interpretation may be even more reliable than human annotation (Wong et al. 2009; Zaidel et al. 2009; Taghva 2010; Moran et al. 2006; Cagnan et al. 2011).

Other researchers have shown that a complex visualization of multiple features derived from the μEEG signal may lead to improved speed and STN identification accuracy (Falkenberg et al. 2006). The Power Spectral density has been shown to be a useful discriminative feature for both visualization and classification (Pesenti et al. 2004; Falkenberg et al. 2006; Novak et al. 2007; Novak et al. 2011), including works published by the author of this thesis (Bakstein 2011; Wild et al. 2010).

A whole branch of research is also focused on evaluating propagation of the stimulatory pulses in the tissue, side effects and optimal stimulation contact placement with regard to these factors (Butson et al. 2007; Bériault et al. 2012) — some researchers have suggested even multiple-contact stimulation electrode, steering the electrical field to the desired area (Pollo et al. 2014).

The DBS surgery is a process with high demands on accuracy and surgical team experience — as such, sophisticated visualization, classification and modeling tools may contribute to accuracy of the procedure — contributions to the field of automatic nuclei classification and modeling can be found in Chapters 5 and 6

2.3 DBS-related signals

In terms of digital signal processing, two types of DBS-related signals will be discussed in this work: the microelectrode recordings (MER), obtained during the surgical procedure, and Local Field Potentials (LFP), recorded through electrodes with larger area — most commonly stimulation electrodes used during the surgery or stimulation electrodes of a stimulator device. Brief characteristics of both signal types are given in this section.

2.3.1 Microelectrode recordings

In connection with DBS, MER (also micro-EEG or μ EEG) signals are very fine recordings of neuronal activity, obtained using microelectrodes with exposed tip size of just tens of micrometers (Slavin et al. 2004). Such electrode dimensions, comparable to dimensions of the neurons themselves, allow capturing an extracellular image of activity of neurons in the closest proximity of the electrode. Such recording contains typically unit activity of one to several neurons, plus background noise, representing activity in areas more distant from the electrode.

Thanks to the high level of detail of the MER, activity (i.e. usually the firing times) of

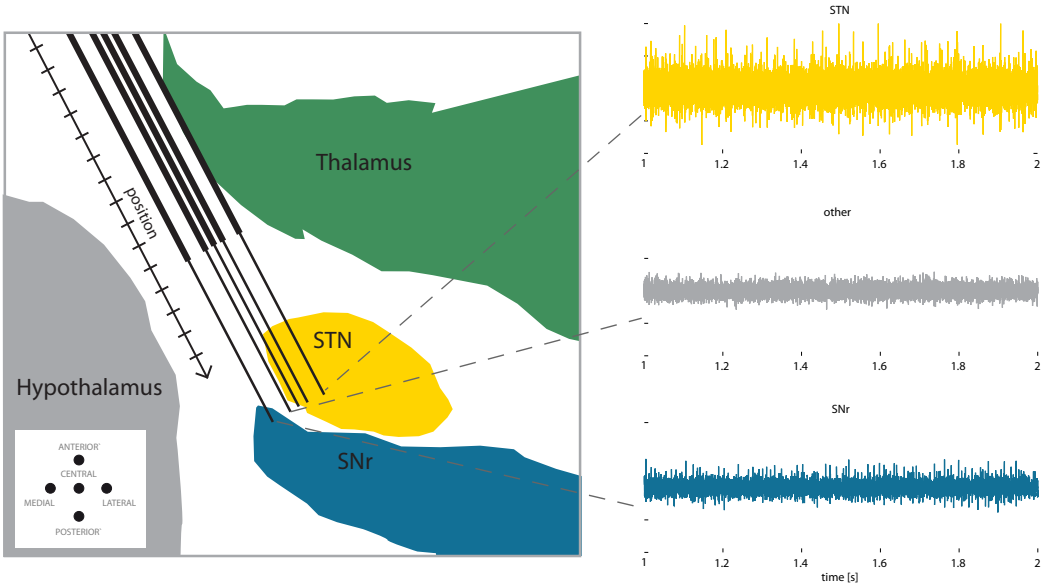

Figure 2.2: Electrophysiological exploration principle: a set of five microelectrodes is shifted through different areas of the basal ganglia, following designated collateral trajectories. The surgical team than identifies STN boundaries from the recorded MER and determines optimal stimulation electrode placement. Compare to exploration protocol in Fig. 2.3.

individual neurons may be studied. To identify activity from close neurons, when only the recorded mixture signal is available, the methods referred to as *spike detection* may be adopted (Quiroga et al. 2004). Sophisticated clustering methods can then be used to attribute individual action potentials to different neurons in electrode vicinity. More information about the spike sorting process can be found in (Quiroga et al. 2004; Wild et al. 2012) and an example of a MER signal with an illustration of signal generation and decomposition can be found in Figure 2.4.

As described above, microelectrode recording (and optionaly microstimulation) is often carried out prior to implantation of the stimulation electrode to refine on the target position. One or multiple (commonly five) microelectrodes in a spaced-away collateral setup are shifted into the brain using motor-driven microdrive and unit activity in their proximity is recorded. Based on the recordings, an experienced neurologist is able to match the observed activity to one of the supposed nuclei and determine spatial margins of the target structure with respect to the electrodes. This process is usually based on visual and auditory inspection of the recorded MER and the results are written down to a protocol (example in Figure 2.3). An illustration of the targeting scheme can be found in Figure 2.2.

One important property of MER signals is the unavailability of absolute signal amplitude.

This is given by two reasons, first of which is blunting of the sharp electrode tip during the shift through the tissue, which affects impedance characteristics of the electrode and therefore the recorded signal. The second reason is inherent in the recording setup and is based on the electric field of a neuron decreases substantially with increasing distance from the electrode. A neuron very close to the electrode will produce signal several times higher in amplitude than neighboring neuron several tens of micrometers further apart (see Mesa et al. 2013 for an explanation of MER generation model). Unlike the electrode blunting that affects both unit activity and background noise, the distance between the neuron and the electrode impacts the unit activity only. It is important to note that the situation of neuron positions around the electrode tip is random due to the dimensions in question and can not be influenced intentionally in clinical setting

The typical properties of a microelectrode recording system make the resulting signals susceptible to movement and electromagnetic artifacts. More details can be found in Chapter 4, where a system for automatic MER artifact identification is presented.

2.3.2 Local field potentials

Contrary to the MER, the LFP signals are commonly obtained using electrodes of a much larger area — typically in the order of units of mm^2. The LFP signals are most commonly recorded in DBS in two settings: i) intra-operatively (the microelectrode shafts usually contain larger electrodes for short-term test stimulation in the assumed target area) or ii) post-operatively using the stimulation electrodes at that time firmly set in the brain. In the latter case, a possibility of LFP recording is in the peri-operative period, when the stimulation electrodes are already in place and the patient stays in hospital for observation before the transient effects subside and the stimulator can be implanted and leads internalized. During this period, the electrode leads are available for recording using external device. This setup provides a possibility to perform various experiments while recording the LFP signals. To monitor concurrent motorical symptoms, surface EMG on patient's forearm (extensor or flexor muscle) is often used. This is especially useful for monitoring of Parkinsonian tremor and this approach has been used in the research presented in Chapter 3.

Another option of LFP signal recording is directly by the stimulator device. Although a majority of devices currently implanted worldwide do not support LFP recording (*Medtronics online catalog* 2012), devices that provide this functionality are undergoing experimental testing and are likely to be introduced to clinical practice in the upcoming years.s. It is very likely that LFP recording and on-board processing will soon become a standard part of the DBS devices, used for automatic stimulation parameter adjustments, inspection of patient state, review of the therapeutic process, and other.

Parameter	MER	LFP
Electrode type	Exploration microelectrodes	(Stimulation) macroelectrodes
Order of contact size	μm	mm
Sampling frequency	12 or 24 kHz	$200 - 1000$ Hz
Activity captured	individual neuron firing	summary activity

Table 2.1: Comparison of typical properties of MER and LFP signals.

The multielectrodes, used commonly for stimulation and LFP recording, are typically fitted with a set of 4 electrode contacts, spaced $0.5 - 1.5$ mm away from each other. In the therapy, only the contacts located in best-suited parts of the target structure are used for stimulation. Due to much larger area of the electrode tips than in the case of MER recordings (order of mm^2), the signals capture summary activity of larger region around the electrode and identification of activity of individual neurons is not possible directly. In this respect, the LFP signals appear closer to classical electro-encephalogram (or EEG) than the MER. Attempts have been made recently to reconstruct some aspects of spiking activity from the LFP signals (Michmizos et al. 2012). However, the techniques are still in development and sensitive to correct parameter setting.

The differences between both signal types are summarized in Table 2.1. Despite the clear distinction used in this text, the dependency of signal parameters on electrode size and projection between MER and LFP signals is still subject to active research (Winestone et al. 2012; Verhagen et al. 2015).

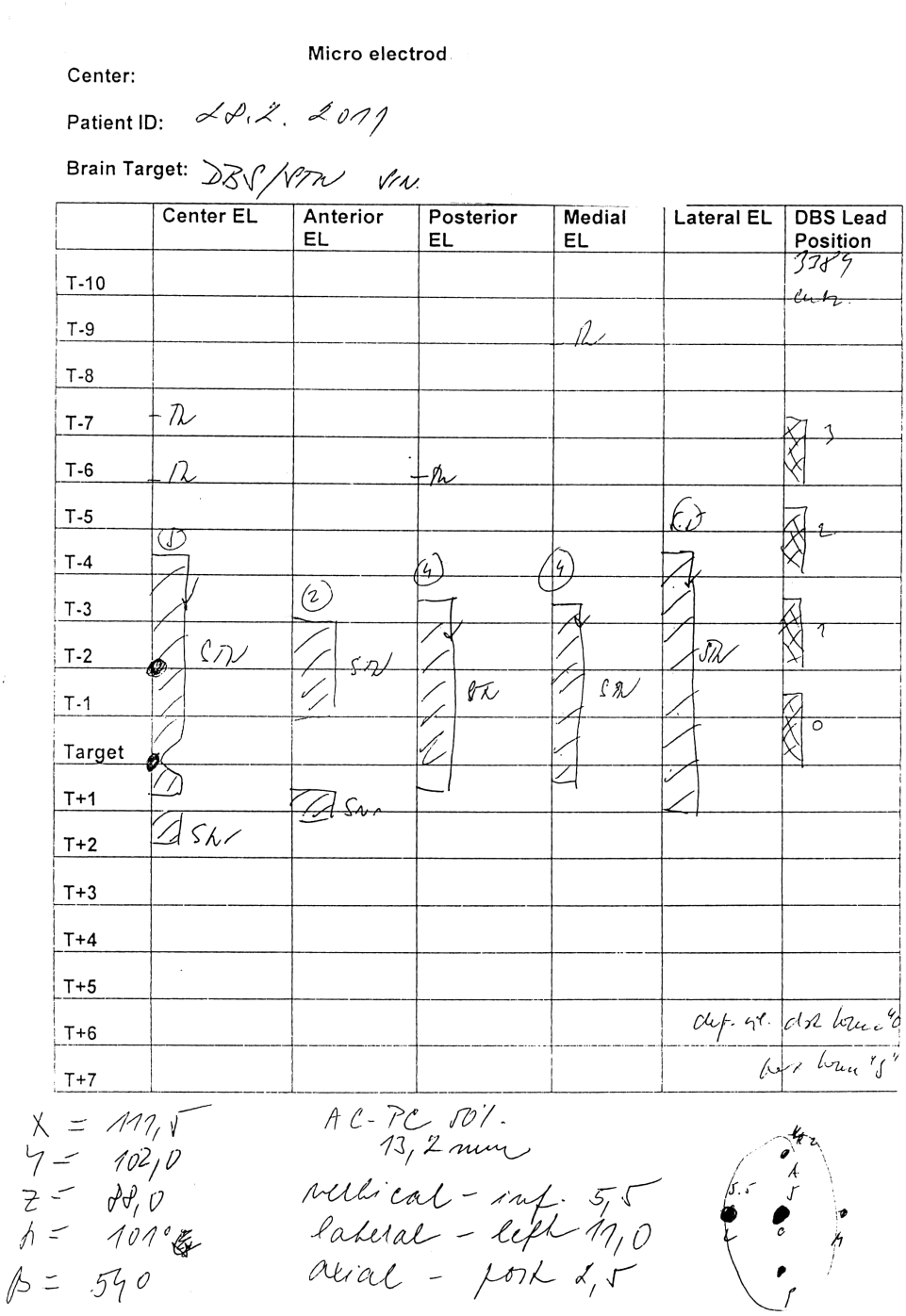

Figure 2.3: Example of clinical protocol from MER exploration: The neurologist identifies the most likely STN passes based on the neurophysiological activity and determines final trajectory and stimulation contact positions (right column). The bottom left part of the protocol shows stereotactic frame settings, the sketch at the bottom right shows situation of STN with respect to the electrodes. Sensitive information has been erased.

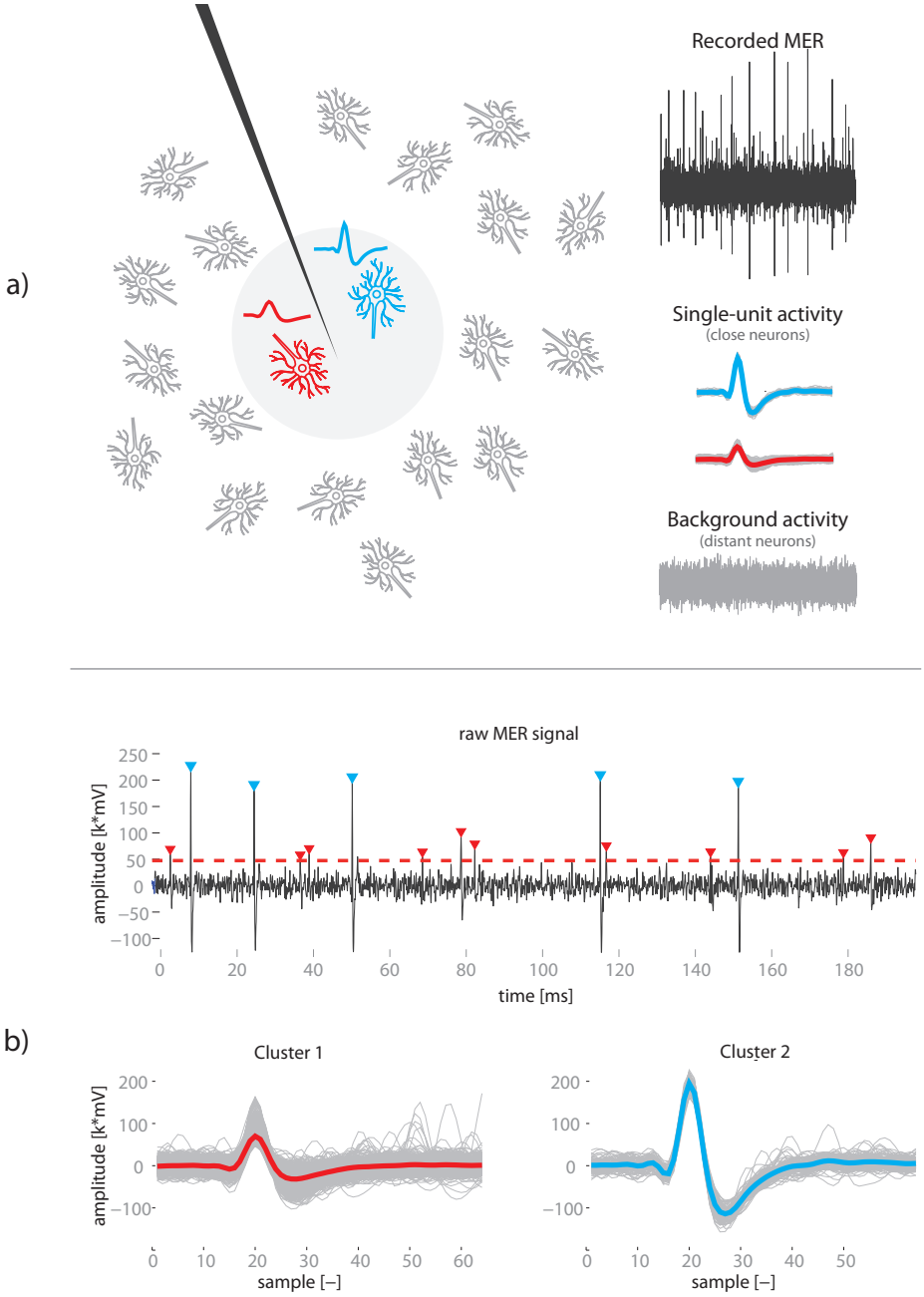

Figure 2.4: a) Components of MER signal: the recorded series is a summary activity of a large number of neurons further away from the electrode (*background activity*) and activity of neurons in close vicinity of the electrode tip (*single-unit activity*) — see e.g. (Martinez et al. 2009) for details. The signal can be decomposed in an estimation procedure — a process called *spike-sorting* b) likely close neurons are first detected using an amplitude threshold (red dashed line). Peaks exceeding the threshold form a set of candidate spikes, which can than be clustered based on their amplitude and shape.

Chapter 3

Tremor onset detection in local field potentials

Chapter summary

One of the scientific topics in DBS-related research is the effort to move from stimulation systems with hard-coded parameters to adaptive, closed-loop systems which adjust stimulation parameters according to current state of the patient (see Section 2.2.1 for introduction). This chapter describes a tremor detection algorithm, operating on LFP signals recorded through the stimulation electrode: a system developed using this approach might reduce DBS side effects and increase battery life by switching the stimulation on only when the tremor is detected. The system is based on an artificial neural network and multiple signal transformations and was developed in cooperation with colleagues from the University of Reading and published as:

- **Bakstein, E.**, Burgess, J., Warwick, K., Ruiz, V., Aziz, T., Stein, J. (2012). Parkinsonian tremor identification with multiple local field potential feature classification. In: *Journal of Neuroscience Methods*, 209(2), 320–330.

This paper also forms the basis of this chapter's text. Some suggestions from our initial research were further investigated in a subsequent paper, to which my contribution was only minor. Several findings from this paper have been added to the final discussion.

- Camara, C., Isasi, P., Warwick, K., Ruiz, V., Aziz, T., Stein, J., **Bakstein, E.** (2015). Resting tremor classification and detection in Parkinson's disease patients. In: *Biomedical Signal Processing and Control*, 16, 88–97.

3.1 Introduction

Parkinson's Disease (PD) is a chronic progressive neurodegenerative movement disorder, which is estimated to affect 6.3 million people worldwide (Baker et al. 2004) and for which there is no recognised cure. The pathophysiology of PD is idiosyncratic in nature resulting in a variety of symptoms displayed by patients, the cardinal clinical symptoms of which are bradykinesia, postural instability, and most notably, rest tremor (Rajput et al. 1997). The frequency of a Parkinsonian tremor does not refer to the firing rate of muscle fibres but the movement of the limb as a whole. This individual frequency is predominantly exhibited from 4 to 9 Hz and in the context of muscular movement it is referred to as the tremor frequency (Carr 2002; Deuschl et al. 1998). Parkinsonian tremor has been reported to be driven by the abnormal activity of neural signals propagated throughout the sensorimotor system (Hammond et al. 2007; Morrison et al. 2008), which we aimed to identify in this study. The recorded LFP signal represents low frequency components originating from axons, somata and dendrites around the electrode, and thus mainly reflects the input to the local brain region. The exact frequency range of the LFP is, in general, below 100 Hz (Brown 2003). However, the majority of neurophysiological studies on Parkinsonian patients have focussed on LFP oscillations in the range of 5 to 35 Hz (Steigerwald et al. 2008; Trottenberg et al. 2007). Current biomarkers for Parkinsonian tremor analyse the oscillations of the LFP signal. The most commonly studied neural activity in the LFP is that of the tremor and beta frequency bands.

It has been found that some groups of STN and GPi neurons display tremor-related bursts with a high coherence to the frequency of spontaneous muscular tremor (Amtage et al. 2008). Such neuron pairs are commonly termed "tremor cells". Synchronized tremor activity is not clearly understood and it is presently unknown as to whether the activity of tremor cells contributes to the development of tremor or is simply an artefact driven by the physical tremor. The presence of tremor cells and the activity of the beta band are however known to be bound to Parkinsonian tremor (Amtage et al. 2008; Zaidel et al. 2010). The results dispel the belief that tremor cells are a manifestation of purely the lack of dopamine, yet still the threat of a physical artefact cannot be ignored. Overall, these investigations suggest the pathophysiology of PD is in close relation to the changes apparent in STN activity.

Apart from tremor-related activity in the STN and GPi nuclei, thalamus and cortex are often related to the mechanism of tremor. Previous studies have shown, that LFP recordings from these sites show correlation with muscular (EMG) activity (Marsden et al. 2000). However, as the majority of DBS implants are nowadays placed either in the STN or GPi nuclei, the focus of this study is aimed at LFP signals from these regions. A number of studies have shown significant coherence between tremor cells and EMG activity in the STN at the tremor frequency (Rodriguez et al. 1998). The tremor frequency and beta frequency have been of great interest in

previous studies (Lemstra et al. 1999; Marceglia et al. 2009), and are reported to be akinetic, i.e. inversely related to motor activity. Some studies discuss the coherence between beta activity and common symptoms of Parkinson's disease, such as akinesia or rigidity. Also, they have been reported to be strongly modulated by voluntary movement activity and medication. The range of the tremor frequency varies between studies. In the context of muscular activity a range of 4–9 Hz is common, however neurophysiological studies analysing LFP signals tend towards lower frequencies (3.0–4.5 Hz (Wang et al. 2007) and 3.0–6.0 Hz (Lemstra et al. 1999)).

The beta-band is a range of frequencies between 12 and 30 Hz and is used in neuroscience to describe the oscillatory range of brain activity. Beta-band synchronisation between neuron pairs in the thalamus has been reported in the majority of investigations into Parkinsonian tremor (Amtage et al. 2008; Levy et al. 2002). It has already been demonstrated that beta activity has a direct impact on post-surgical prognosis, i.e. electrophysiological factors bearing direct relevance to surgical outcome (Marsden et al. 2001; Ray et al. 2009b; Zaidel et al. 2010). This may in the future be detected pre-operatively through non-invasive imaging techniques, e.g. (Ray et al. 2009a). Further, high-frequency stimulation has been shown to suppress subcortical beta activity (Kuhn et al. 2008). Activity in the beta range of the STN has also been found to be a strong marker for the sensory-motor region of the STN and was successfully used to identify this area in Microelectrode recordings (Zaidel et al. 2009).

Wang et al. reported on a coherent relationship between the onset of PD tremor and LFP oscillation activity, in both tremor (3.0–4.5 Hz) and beta-band (10–30 Hz) frequencies (Wang et al. 2005). Coherence was evident between LFP and EMG signals at the tremor frequency, whereas the power exhibited at the beta-band frequencies decreased prior to and throughout tremor activity.

Given the idiosyncratic nature of the disease we propose here that to achieve a high rate of tremor classification over a broad range of patients, a concise biomarker will benefit from exploring more than just the frequency feature of the LFP. To improve upon current biomarkers, we therefore explore here a multi-feature classification approach to identify Parkinsonian tremor within an LFP signal.

Selective properties of individual features from the temporal and spectral domain were tested in the study with optimal properties being used for the classification of tremor. Moreover, the suitability and relevance of the different approaches was compared. The classification results were then used to assess the actual applicability of the proposed method. As a result of this, features suitable for tremor detection can be differentiated from features with little or no significance to the problem. The results obtained in this study further our understanding of Parkinsonian tremor and ultimately will enhance the maintenance of personal health.

3.2 Materials and methods

In this section we have summarized the range of experimental design procedures carried out in terms of recording signals, conditioning signals and evaluating the classification of system performance.

The overall task requirements were as follows: Firstly simultaneous LFP and EMG recordings were made using volunteer patients with implanted deep brain electrodes. Secondly the data obtained — particularly deep brain LFPs — were divided into smaller chunks (windows) and different features were measured, these included temporal features, spectral features and features founded on information theory. Thirdly feedforward Neural Networks were trained using only a subset of data based on the leave one out philosophy, the one left out being the patient whose data would be subsequently analysed. This made it as difficult as possible for the overall analysis to investigate each patient as it had to do so without seeing any data from that particular patient apriori. Apriori viewing, and even training on a particular patient's data would have made the subsequent results appear to be far better than those actually obtained – in the presented case the subsequent results should be seen as 'worst case' outputs.

As each of the features are introduced, reasons are given, based both on previous research and subsequent analysis as to why each feature was considered. In any case, in the results section it will be seen that any of the features which didn't actually turn out to be particularly useful were effectively dropped. The list here should therefore seen to be one which includes all reasonable possibilities. For the neural networks and signal transformations, concepts of their employment are described.

3.2.1 Subjects and data acquisition

Eight volunteering patients, who had been diagnosed with tremor-dominant idiopathic PD, participated in this study as they were undergoing surgical DBS treatment at the John Radcliffe Hospital (JRH), Oxford, UK at the time. Postoperative magnetic resonance imaging confirmed the successful implantation of the DBS electrode lead(s) (Medtronics 3387) within the target region.

We exploited the external bi-directional connection to the electrodes, which was available during the intra-operative period of stereotactic surgery, to record neural and coherent muscular signals from all of the volunteering patients. Approval for the recording sessions was granted by the local research ethics committee of the Oxfordshire Health Authority, UK, and informed consent was obtained from each volunteer.

The DBS electrode lead employed was Medtronic 3387, with four electrodes spaced 1.5 mm apart which permits three different contact pair configurations (0+1, 1+2, and 2+3) to be simultaneously recorded in a bipolar fashion with one contact used as reference. Muscular

patient ID	target	files	total length [s]	Chunk counts			
				Atrem.	Onset	Trem.	Total
P1	STN	1	60.0	114	30	146	290
P2	STN	4	147.0	253	120	325	698
P3	STN	1	60.0	101	30	159	290
P4	GPi	1	25.0	51	30	35	116
P5	STN	1	30.0	38	30	72	140
P6	GPi	2	97.0	205	60	201	466
P7	STN	2	48.5	38	60	126	224
P8	STN	1	60.0	128	30	132	290
Total		13	527.6	928	390	1196	2514

Table 3.1: Dataset overview. Chunks produced using 2-secs windows with 90% overlap.

(EMG) signals were simultaneously recorded using disposable adhesive Ag/AgCl electrodes (H27P, Kendall-LTP, MA, USA) placed in a tripolar configuration (active–common–reference) over the tremulous forearm extensor (carpi ulnaris) and flexor (carpi radialis).

The measured signals were then amplified using isolated CED 1902 amplifiers ($\times$10000 for LFPs and $\times$1000 for EMGs), filtered at 0.5–500 Hz and digitized using CED 1401 mark II at rates between 250–1000 Hz. For the patients concerned, after a two week recovery period the electrode leads were internalised and connected to an implanted pulse generator (Medtronic KINETRA model 7428) which was surgically inserted into the chest cavity.

3.2.2 Dataset description

Our entire database of recorded raw LFP signals consisted of 13 LFP tremor onset recordings from a total of eight patients – the details of which are given in Table 1. Out of the 8 patients included in the study, only two — P1 and P2 — were implanted in the GPi nucleus, while the remaining 6 received STN implants. The data chunk counts presented in the table were obtained using 2 seconds long chunks (500 samples) with 90% overlap. Onset chunks were those found in a 4 seconds region around EMG-based tremor onset (see the next section for further details). Note: Each chunk consists of collected data within a 2 seconds window.

The maximum amount of recordings taken from a single patient was four, although the median was only one recording per patient (Table 1). This was caused mainly by the requirement for tremor onset to be present in the recording period. The total number of recordings was 13, the total length of time 527.6 secs, and the average length of each recording was 40.5 secs.

3.2.3 Data preprocessing

Firstly, to maintain a uniform frequency across recordings, all data were down-sampled to 250 Hz (the lowest sampling frequency used). A 3–30 Hz and 3–7 Hz (tremor band) Chebyshev Type II passband filter was used on the recorded LFP and EMG signals respectively. The EMG signal was then normalized and rectified. Low frequency LFP activity systematically relates to slow drift movement artefacts, while frequencies above the beta-band (30 Hz) are considered to hold little tremor-related information. Indeed this frequency range also includes abundant 50 Hz line noise.

Tremor onset was calculated for each file based on the amplitude of filtered and rectified EMG signal. The magnitude of the EMG time series value was checked against a threshold of three times the mean of the EMG magnitude in the first 5secs of the recording (known to be recorded during atremorous activity — that is LFP activity when no tremors were apparent). If an indication of high tremor frequency activity was determined at any point in time then the mean of the next 5secs of data (long enough to cover any period of small tremor-onset) was calculated to confirm the initial tremor detection. A single time of tremor onset was calculated for each patient recording. This calculated point was then used for data annotation and division of the data into into tremor and atremorous sets.

3.2.4 Feature extraction

Feature extraction involves representing the raw signal in terms of a smaller set of quantities, termed features. Feature measurement from biological signals is a routine process in biomedical studies, however, correct feature selection is fundamental in order to relate signal characteristics to various biological measures such as the detection and classification of tremor in an LFP signal. Many features or feature sets exist in biomedical engineering for signals such as ECG, EEG and EMG (Ciaccio et al. 1994), whereas LFP signals taken from humans have only been readily available to researchers since the early 1990's.

In this section we consider a range of different features and explain how their characteristics are obtained from the original raw LFP data. The LFP data here is divided into two second chunks with 90% overlapping so a spatial resolution of 0.117 Hz and a time resolution of 0.2 secs was achieved. Evaluation of the performance of individual features in different patients is given in section 3.3.1.

Time domain

Increased LFP oscillation energy is known to be related to the excitation and synchronised firing of active neurons in the motor cortex (Denker et al. 2007) and in the Parkinsonian STN

(Loukas et al. 2004; Rivlin-Etzion et al. 2006). Furthermore, discharges of single STN neurons are often coupled to Parkinsonian tremor oscillatory activity (Kuhn et al. 2005; Moran et al. 2008; Weinberger et al. 2006).

LFP activity from the basal ganglia is a signal of infinite length (a lifetime) and so it is impossible to assert the "true" variance of the population (all neurophysiological data). Instead a sample variance can be measured from the finite atremorous and tremorous LFP signals provided. For this an unbiased variance calculation is used to compensate for the discrepancy of the sample and population. Given that LFP signals oscillate in time, the sample variance is never exactly zero, consequently, the larger the variance, so the more variability there is in the measured sample.

The variance feature (termed *var*) is designed to simply return the variance of the given LFP signal window and it is used to analyse abnormal LFP oscillatory activity. The magnitude of the *var* feature is hypothesised to increase during periods of tremor, at which times sporadic abnormal neuronal activity is known to develop. The sample mean, $\overline{x}$, and sample variance, var, of a signal can be estimated by

$$\overline{x} = \frac{1}{N} \sum_{t=1}^{N} x(t) \tag{3.1}$$

for the sample mean and

$$var(x) = \frac{1}{N} \sum_{t=1}^{N} (x(t) - \overline{x})^2 \tag{3.2}$$

for sample variance of the time series $x(t)$ of size N.

The zero-crossing rate (a feature name *zerox*) is arguably the simplest form of frequency analysis in the time domain. This feature crudely measures the frequency of a signal with zero mean by counting the number of times the LFP signal value crosses zero magnitude in a given time window. Consequently, higher frequency signals produce a greater zero-crossing rate and vice versa.

Quite a number of zero crossing based frequency estimation methods have traditionally been used in applications such as speech and communications (Anderson 1982; Wiley et al. 1977), additionally a study of the zero crossing rate has also been used in biomedical engineering for EMG (Masuda et al. 1982; Skotte et al. 2005) and EEG (Wua et al. 2011) analysis. This feature counts the number of times the LFP signal value crosses zero magnitude in a segment of LFP data, and can be expressed as

$$zerox(x) = \sum_{t=2}^{N} \begin{cases} 1, & \text{if } x(t) = 0 \ OR \ x(t)x(t-1) < 0 \\ 0, & \text{otherwise} \end{cases} \tag{3.3}$$

where N is the length of the segment.

The application of autocorrelation in neurophysiological signals was first used to analyse EEG signals as early as 1950 (Brazier et al. 1952), and despite the introduction of the FFT, the autocorrelation function is to date commonly employed to estimate statistical moments of the PSD (Derya Ubeyli 2009; Jackson et al. 2008; Tagluk et al. 2011). Autocorrelation is useful for finding repeating patterns in a signal, determining the presence of a periodic signal which has been buried under noise, and identifying the fundamental frequency of a signal which doesn't actually contain that frequency component, but implies it with many harmonic frequencies.

Autocorrelation is simply the correlation of a discrete time series $x(t)$ against a time-shifted version of itself. The discrete autocorrelation R at time lag j for the signal $x(t)$ is defined as

$$R(j) = \sum_{t=1}^{N} (x(t) - \overline{x})\,(x(t-j) - \overline{x}) \tag{3.4}$$

where $\overline{x}$ is the sample mean

The area under the curve between bands is calculated and used for the *ac_tremor* and *ac_beta* features, written as

$$ac\_band = \sum_{j=\beta_2}^{\beta_1} |R(j)| \tag{3.5}$$

where β_1 and β_2 are boundaries of the frequency band in units of lags, given by

$$\beta_k = \frac{N}{f_k} \tag{3.6}$$

In which N is the length of the signal and f_k is the pseudo frequency in Hz.

Information theory

All the features presented thus far share a common goal; they are all used to try to differentiate information in the LFP signal between the two main Parkinsonian tremor relationships (tremor and atremorous). The direct existence of Parkinsonian tremor relationships can however be further analysed by means of information theory. Central to this theory is the concept of entropy, which can be roughly defined as a measure of the uncertainty of a signal. A known example of a Probability Density Function (PDF) on a set of real numbers is the Gaussian distribution or normal distribution, and this can be used to describe synaptic potential amplitudes.

Because the LFP signals collected are not of Gaussian distribution (each patient recording failed the one-sample Kolmogorov Smirnov test at $p < 0.05$), the entropy is calculated using five different PDFs: i) A PDF derived from a histogram of the entire tremor-onset signal (this feature is named *entropy_entire*). ii) A PDF derived from a histogram of the atremorous LFP signal of a patient recording makes the feature *entropy_atrem*. iii) A PDF derived from a histogram

of the tremor LFP signal of the patient recording makes the feature *entropy_trem*. vi) A PDF derived from a histogram of the windowed LFP data termed *entropy_win*. v) Finally, despite the failed Kolmogorov Smirnov test a normal PDF is created using the mean and standard deviation of the entire tremor-onset signal, this feature is named *entropy_norm*. The feature output is given by

$$entropy(p_k) = -\sum_{t=1}^{N} p_k\left(x(t)\right) \log p_k\left(x(t)\right) \tag{3.7}$$

where k is the data used to calculate the PDF p and this therefore defines the entropy feature.

Frequency domain

Pre-defined frequency bands classifying activity in the human brain can be used to determine its current functional state, and changes in the spectral properties of LFP activity recorded from the basal ganglia of a Parkinsonian patient are also known to correspond to tremorous movement (Amtage et al. 2008). The spectrum activity of the basal ganglia can be classified into four groups or bands: tremor (3.5–5.5 Hz), theta (5.5–8 Hz), alpha (8–12 Hz), and beta (12–30 Hz). The gamma band (30–100 Hz) was contaminated with heavy 50 Hz noise and was thus removed in the pre-processing stage, as described earlier in section 3.2.3.

The discrete Fourier transform is a function that decomposes a signal into its constituent frequencies, and can be obtained by

$$F(u) = \sum_{t=0}^{N-1} x(t) e^{-2\pi i u t/N} \tag{3.8}$$

The $freq\_band$ feature looks at the frequency properties of the signal and calculates the magnitude of the PSD within the boundaries of alternate frequency bands, here

$$freq\_band = \sum_{u=\delta_1}^{\delta_2} |F(u)|^2 \tag{3.9}$$

where $F(u)$ is the discrete Fourier transform of discrete time signal $x(t)$, and δ is the boundary of the frequency band. Figure 3.1 shows spectral comparison of averaged atremorous and tremorous LFP signals.

The tremor frequency band used in this feature is set as 3.5–5.5 Hz. This band was selected in particular because it matches the appearance of increased PSD during episodes of tremor, as can be seen in Figure 1: Spectral comparison of averaged atremorous and tremorous LFP signals. The theta-band (5.5–8 Hz) is though closely related to the tremor-band, and may

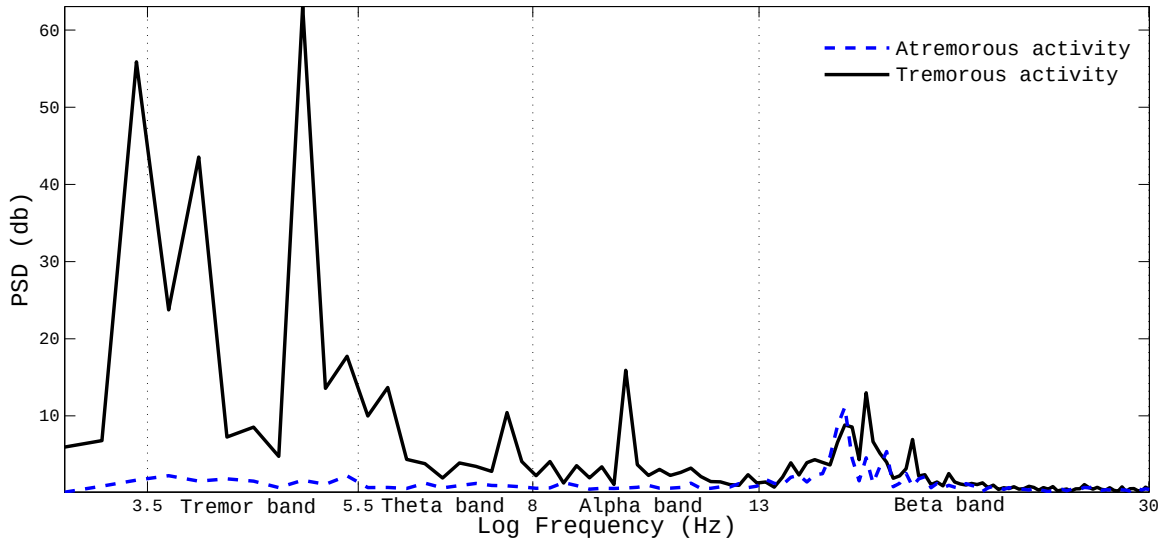

Figure 3.1: Spectral comparison of averaged atremorous and tremorous LFP signals

capture some outliers. In such an event the PSD mean from tremor periods is expected to follow closely to that of the tremor-band and be slightly higher than from atremorous periods, albeit not as significant. On the other hand, the alpha-band (8–12 Hz) has not been reported in previous research to have a significant association with PD tremor.

It is noticeable that in Figure 3.1 we have selected a window 3.5–5.5 Hz, yet it can be seen that there is significant activity below this frequency band. We have selected the frequency range in question because of the desire to select one range to cover the published tremor frequency range (Hutchison et al. 1997; Lemstra et al. 1999; Liu et al. 2002; Magarinos-Ascone et al. 2000; Wang et al. 2007). We believe that it will make an interesting future study to consider a two window method, splitting the tremor range between the to two frequency bands. An option to build a robust tremor-detecting feature could also involve searching for a maximum (tremor-frequency peak) in the tremor range. However, this inspection would require further evaluation of the tremor-frequency signal in the LFP signals, which can be verified in future studies.

Continuous Wavelet Transform

The Continuous Wavelet Transform (CWT) is similar to the time-frequency method of the Short Time Fourier Transform (STFT), but rather a time-scale analysis is performed. The compromise between time and frequency information from an STFT can be practicable, nonetheless a fixed length window is used throughout the process. Nonstationary signals like LFPs can often require a variable-sized window approach which can enhance either time or frequency information.

The length of the window determines either precise low-frequency information using long segments, or high-frequency information using shorter segments. Unlike the Fourier transform, which breaks up a signal into sinusoids of various frequencies, the CWT decomposes a signal using a family of 'wavelets' varying in scale to collect shifted and scaled information. An estimation of frequency, called the pseudo-frequency, can be calculated from the scale-time signal using Equation 3.10, and this is employed to measure the activity in the tremor- and beta-bands.

$$a = \frac{F_s F_c}{F_a}, \tag{3.10}$$

where a is the scale, F_s is the sampling frequency, F_c is the centre frequency of the wavelet and F_a is the pseudo-frequency[1].

The usual notation of the Continuous Wavelet Transform (CWT) is

$$CWT_x(a,\tau) = \frac{1}{\sqrt{a}} \int x(t)\varphi\left(\frac{t-\tau}{a}\right) dt \tag{3.11}$$

In which φ is the basic wavelet function or the so-called *mother wavelet*, a is the scale factor and τ is the translation in time.

In the CWT feature we chose the Shannon wavelet as the mother wavelet. A Shannon complex wavelet is symmetrical in shape and can compute the complex continuous wavelet transform when given two adjustable parameters, the bandwidth and centre frequency. Using a bandwidth of 0.2 and centre frequency of 2 Hz, a good temporal-spectral resolution is achieved for analysing the tremor frequency (3.5-5.5 Hz) and the beta frequency (12-30 Hz) in LFP signals from the patient database. This feature is named *wav_#*, where # is the frequency band (either *tremor* or *beta*).

Discrete Wavelet Transform

The Discrete Wavelet Transform (DWT) is an extension of the CWT and is a generalization of wavelet decomposition that offers a rich range of possibilities for signal analysis. The transform decomposes a signal into an approximation and a detail. The approximation or detail itself is then split into a second-level approximation and detail, and the process is repeated until the desired level of decomposition is reached, resulting in a complete binary tree of all possible decompositions. The approximation of a signal is obtained by convolving the signal with a low-pass filter and with a high-pass filter for detail decomposition.

The DWT feature termed wavdetail returns the detail wavelet decomposition at the 5^th level of the approximation. After which frequencies in the high beta-band are suppressed; leaving

[1]The pseudo-frequency is only an approximation of the frequency because the centre frequency is an approximation of the dominant frequency in the wavelet oscillations

the tremor, theta, alpha, and low beta-band frequencies in the signal. The DWT of a signal $x(t)$ is calculated by passing it through a series of filters:

$$DWT_{high}(u) = \sum_{t=1}^{N} x(t)g(t-u) \tag{3.12}$$

$$DWT_{low}(u) = \sum_{t=1}^{N} x(t)h(t-u) \tag{3.13}$$

where g is a low-pass filter which returns approximation coefficients, and h is a high pass filter which returns detail coefficients.

After each DWT transform (or decomposition) the coefficients are down-sampled by two, since half the frequencies of the signal have been removed in the process. The feature wavdetail is the sum of the squared magnitude of the 5^{th} detail of the signal $x(t)$, written as

$$wavdetail = \sum_{u=1}^{N/2^5} |DWT_{high}(u_4)| \tag{3.14}$$

where u_4 is the 4^{th} detail decomposition of $x(t)$.

3.2.5 Neural network multi-feature classification

In a multi-feature classification, the discriminative power of the classifier is based on a set of observed or calculated characteristics of the training examples — the features. The main advantage of this approach is that even features showing low discriminative power when observed individually (a single-feature classification) can potentially achieve better classification performance when combined together. This is based on the multi-dimensional information, contained in the whole features set, assuming low inter-feature correlation.

In this study, we assumed that the tremor-related changes in the LFP signals were not bound exclusively to one type of signal transformation — such as frequency spectrum — but that these changes could be observed from different signal properties. To accomplish this, features based on different properties of the LFP signals were combined in a single classifier, which should lead to the utilization of more tremor-related information.

Feature selection

To design a classification system with good performance, a set of highly relevant features is necessary. Even though the features calculated according to the description in section 3.2.4 were designed to capture observed changes in the LFP signals, their actual classification power for the problem considered was unknown until they were applied to the data in question.

To evaluate the tremor classification power of individual features, a relative information gain metric was adopted. The relative information gain value represents the relative drop in entropy of the system in the case when a particular feature's value was known. In the simplified case of two random variables, the relative information gain (denoted by G) can be represented by the formula given as:

$$G(X) = \frac{H(Y) - H(Y|X)}{H(Y)} \tag{3.15}$$

where $G(X)$ is the relative information gain of feature X, $H(Y)$ is the entropy of attribute Y and $H(Y|X)$ is the conditional entropy of feature Y, given that the value of X is known.

Before the features were entered into the input layer of the neural network model, the relative information gain was calculated for each of them. Features with $G < 0.1$ were considered to be suitably insignificant and were, on this basis, excluded from further processing. Results of the feature evaluation and selection are described in section 3.3.1.

Neural Network Classifier

To distinguish tremor and atremorous data based on the feature values, an artificial neural network (or commonly referred to merely as neural network – NN) classifier was used in this study. NN models have universal nonlinear modelling and classification capabilities which together with high noise tolerance makes them suitable for a wide array of diagnostic machine learning tasks (Hornik et al. 1989; Reggia et al. 1993). Due to their listed versatile properties, NN models have been used for the processing of various biomedical signals since the early 1990's (Reggia et al. 1993) and have become quite commonly applied as an analysis technique for biomedical signals such as EEG in the last few years (Ubeyli 2008).

In this study, a feed-forward neural network classifier with one hidden layer was used and trained on the training data subset. The training process was limited to 200 learning cycles in order to avoid model overfitting, which was apparent at higher training cycle counts. The input to the classifier was formed by feature values, calculated on individual 2-secs chunks. The number of hidden layer neurons (neurons in the middle — not input or output layer) was 10, chosen according to the best performance achieved at this network setup. The number of input layer neurons was adjusted in each cycle according to the actual number of attributes after feature selection.

3.2.6 Machine learning procedure

Conventionally a multi-feature classification process consists of building a feature set, classifier training and then evaluating the overall model's performance on unseen data. This section de-

scribes all the aforementioned steps including modifications, done to accommodate the specifics of the tremor detection task.

Feature extraction from raw data was performed in MathWorks MATLAB environment, while the feature selection and classifier evaluation process was carried out in Rapid-i Rapid-Miner data mining software.

Global and patient-specific classification

Before the classification procedure could be embarked upon, consideration had to be given to the problem of handling data from different patients. A basic supposition, necessary for the multi-feature classification system to be feasible at all, is that different PD patients not only share similar characteristics in their brain activity during both tremor and atremorous periods but also exhibit similar characteristics during periods of change from one state to the other.

On this assumption, it is sensible to suppose that the changes in different patients can be described by similar features (e.g. changes in signal power, changes in frequency spectrum etc.). However, it is very likely that LFP signal properties vary dependant on many factors, including individual patient attributes such as sex, age, PD progression and the exact positioning of the stimulation electrodes within the target structure. Thus, the classification system has to be designed in an attempt to address these issues properly.

Generally, there are two possible approaches to the classification of such signals: a) create a model on a per-patient basis, i.e. the model is trained specifically for each patient b) train a global universal classifier, for the detection of any patient's LFP signals. The latter approach puts much higher requirements on the consistency of tremor-related changes across different patients. Despite this, we considered/hypothesized such a system to be feasible and selected a general classifier as the approach to be used in this study.

In a sense we were thinking about the long term aspects of the end product from our research. Potential advantages being an easier setup procedure with no need for physician training together with versatility of use. Moreover, a design process based on such a system can itself bring much new information about patient-dependent changes.

Classifier training and validation

A common measure of classifier quality is classification performance, given typically by accuracy — the ratio of correctly classified examples to the number of all examples — or error (1 - accuracy). To estimate classification performance of the trained model properly, the examples, presented to the model in the testing period have to be unseen by the classifier during training. This approach, contrary to performance on the testing data, gives a good estimate of the generalization properties of the classifier.

The simplest method to calculate performance estimate is split validation, where data is randomly split into two data sets: the training and testing set. The split operation assumes complete separation and independence between the training and testing sets. Due to inter-chunk dependencies, caused by the high overlap between different chunks, this technique was inapplicable to the presented problem. The dependencies were verified by a test in which the data was divided randomly into training and testing sets in the ratio 7:3 and the testing accuracy reached up to 99.75% as opposed to much higher error rates seen at different dataset splitting methods.

To ensure separated training and testing data, the data splitting method used was based on an iterative per-patient approach: in each cycle, data from one of the patients was withheld for testing and the whole training process — feature selection and model training — was performed on the remaining data. The process was repeated with each patient in the position of testing subject. Two accuracy estimates were calculated: per-patient accuracy, and overall accuracy, calculated from the total number of erroneously classified chunks. Onset chunks were left out from the training set, whereas the test set comprised all data of the testing patient, including onset chunks and one to several recordings. The overall process diagram is shown in Figure 3.2.

Apart from avoiding problems based on inter-chunk dependencies, this approach also provides a good idea of how well the system could work in real deployment. In such a scenario, the detection algorithm would be designed and tested on a data sample similar to that used in this study, with no need for patient-specific parameter tuning or other adjustments.

3.3 Results and discussion

In this section we present results on our analysis of the features selected, on our testing of the accuracy of the proposed classification system in terms of its performance and the use of an in silico model to approach the possibility of real-time analysis by this means. It is worth pointing out that, in each case, the system employed was trained up on data from all the other patients in the study before being tested on each individual patient — the system therefore had not actually seen any data from the patient on which it was due to make its analysis.

As will be seen, the study provided a mixed bag of results. In some cases (for some patients) the results were exceptionally good, much better than we could have hoped for apriori. Such results indicate clearly that the approach taken can work very well. On the other hand, for some patients the results were surprisingly poor. This leads one to believe that the approach taken here is merely providing part of an overall, applicable solution.

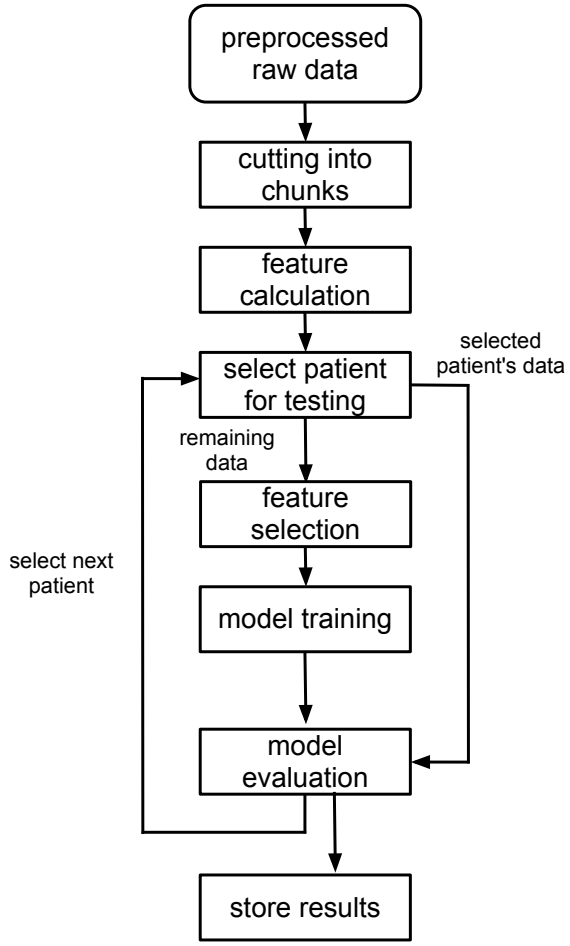

Figure 3.2: Diagram of the leave-patient-out process

Patient	P1	P2	P3	P4	P5	P6	P7	P8	**mean**	**std**
freq_theta	0.69	0.65	0.93	0.97	0.93	0.83	0.49	0.07	**0.70**	**0.27**
wavdetail	0.48	0.61	0.91	0.97	0.93	0.98	0.55	0.04	**0.68**	**0.29**
freq_trem	0.48	0.55	0.96	0.97	0.93	0.74	0.53	0.11	**0.66**	**0.27**
var	0.57	0.26	0.96	0.97	0.93	1.00	0.49	0.06	**0.66**	**0.32**
entropy_norm	0.54	0.27	0.96	0.97	0.93	1.00	0.49	0.04	**0.65**	**0.33**
wav_tremor	0.55	0.26	0.96	0.97	0.93	0.70	0.61	0.12	**0.64**	**0.28**
freq_alpha	0.42	0.10	0.96	0.97	0.93	1.00	0.49	0.14	**0.63**	**0.34**
entropy_atrem	0.52	0.17	0.96	0.97	0.93	0.58	0.78	0.09	**0.63**	**0.31**
entropy_entire	0.50	0.29	0.96	0.97	0.93	0.73	0.29	0.07	**0.59**	**0.31**
wav_beta	0.09	0.06	0.96	0.97	0.93	1.00	0.24	0.20	**0.56**	**0.39**
freq_beta	0.08	0.01	0.88	0.97	0.93	0.46	0.28	0.24	**0.48**	**0.35**
zerox	0.46	0.27	0.93	0.21	0.93	0.13	0.16	0.06	**0.39**	**0.31**
ac_tremor	0.27	0.07	0.75	0.14	0.93	0.24	0.20	0.03	**0.33**	**0.29**
ac_beta	0.58	0.09	0.32	0.05	0.83	0.12	0.21	0.04	**0.28**	**0.25**
entropy_trem	0.11	0.10	0.62	0.36	0.22	0.38	0.09	0.09	**0.24**	**0.17**
entropy_window	0.13	0.04	0.28	0.07	0.37	0.07	0.09	0.03	**0.14**	**0.11**

Table 3.2: Information gain of individual features in different patients

3.3.1 Feature evaluation

According to the procedures described in section 3.2.5, the relative information gain was calculated (minimum = 0.0, maximum = 1.0) for each feature on each patient's data. Chunks from the onset period were excluded from this calculation. In this way, the performance of each feature could be evaluated. The comparative results from each of the different features are presented in Table 3.2 in a relatively ordered fashion.

As seen from the values in the table, despite the fact that the performance of all features varies substantially between patients, the features can be categorized according to overall performance across the set. The first five features in the table – *freq_theta* to *entropy_norm* – are best rated according to their IG value and seem to be the most stable ones according to their top (or at least high) ranking in most patients. The rest of the features showed either mixed performance (e.g. *wav_beta*, *freq_beta*, *entropy_entire*) or poor performance overall (*entropy_trem*, *entropy_window*). What is not perhaps so clear to see from Table 3.2 is that patients' results seemed to cluster and can be divided into the following three groups, based on common IG values: a) patients with IG close to one in most features. This includes patients: P3, P4, P5, P6. Tremor activity in these patients should be possible to classify by means of individual features or small feature groups. b) Patients with moderate IG values. This includes patients P1, P2 and P7. c) Patient P8 with extremely low IG values in all features. Classification of this patient, based on the calculated features seems unfeasible.

patient	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**total**
target	STN	STN	STN	GPi	STN	GPi	STN	STN	
atrem	68,4	94,1	100	100	100	100	71,1	18	**83,1**
trem	92,5	59,4	74,8	100	98,6	71,9	81,8	99,2	**78,1**
total	**81,9**	**74,6**	**90,4**	**100**	**99,1**	**86,2**	**79,3**	**59,2**	**80,4**

Table 3.3: Classification accuracy in individual patients: tremorous, atremorous chunks and overall results. Accuracies calculated from actual numbers of correctly classified chunks — onset chunks not considered.

3.3.2 In silico model

Due to the time needed to record a reasonable amount of data and the fact that the perioperative period, in which the data can be collected, is only a few days long, we were unable to evaluate performance of the system in day-to-day, living conditions. Thus, In order to provide an easy to assess performance test of the whole multi-feature neural network system, we carried out an in-silico test. To split the training and testing data in a manner comparable to real deployment, we evaluated the system using the leave-one patient-out method, where one patient's data is held out for testing and the NN model is trained on the remaining part of the dataset, which means that no data from the testing patient were seen by the classifier prior to testing. The evaluation system was described in greater detail earlier in Section 3.2.5.

Once the data was fully classified – each patient's data having been in the position of a testing subject – all chunks were labeled by the classifier and the performance of the system could be evaluated. To show the results from different perspectives, we have provided here two different types of performance evaluation: common classification accuracy measures and temporal output figures, which will be discussed further, later on.

In Table 3.3 the figures indicate test classification accuracy, calculated as a percentage of correctly classified chunks for the patient in the position of testing subject. Percentages are given separately for atremorous and tremorous chunks. The overall accuracy is calculated from the total number of correctly classified chunks. Onset chunks were not considered in the accuracy calculation.

As seen from the table, quite large differences in classification accuracy are present in the different patients. In two patients – P4 and P5 – the classification accuracy reached almost 100% for both tremorous and atremorous activity, which is outstanding, considering that only data from the other patients were used for classifier training. In another patient group however, including P6, P3 and P2, the classification of atremorous data was very accurate, whilst the tremor detecting capability was low. The completely opposite situation can though be seen in P8, where tremorous data was classified with 99.2% accuracy, while atremorous data was hardly

recognised with only 19%, which means in reality that most chunks were classified as tremorous, even when they were not. Regarding different stimulation – and thus recording – targets, it can be noted that the two patients with electrodes implanted in the GPi were among the patients with the best classification results. However, due to the low number of test examples, no strong conclusion about suitability of either nucleus for tremor detection can be drawn.

A more easily readable representation of the classification results may be obtained in terms of visualization of the classifier output, compared to actual LFP signals and forearm EMG activity. Selected plots, showing these three quantities for three different patients are shown in Figure 3.3.

The classifier output and EMG signal were scaled to match common amplitudes of the LFP signals in order to be easily readable in the plots. Outputs of three patients were selected, representing very good, medium and poor performance respectively. The plots can be matched to corresponding accuracy values in Table 3.3. to gain an overall view of system performance. Unlike in the case of the calculation of accuracy, classified tremor onset chunks were included in these plots to maintain continuity of the system output throughout each file. Discussion of the experimental results is given in the following section.

3.3.3 Discussion of the results

In this study we inspected a hypothesis that the onset of Parkinsonian tremor could be detected directly from the stimulation electrodes and used for on-demand stimulation. Moreover, we suggested that a single parameter setting could be sufficient and tested this assumption on a set of 8 patients. This section aims at commenting on and a discussion of possible causes that lead to the mixed classification results.

The results presented in the previous section show very mixed classification performance in different patients. Comparing classification accuracies in Table 3.3 to feature information gain values in Table 3.2, a link between poor performance of the best features and poor classification accuracy seems obvious (a similar link can be seen in "well performing patients", too). This was probably also the reason for the poor classification accuracy of patient 8: for example features, such as $freq\_theta$ or $wavdetail$, very strong in other patients, showed extremely poor classification power with respect to tremor. Drawing a conclusion about the link between feature performance and classification accuracy seems sensible, however the reasons for poor feature power in some patients are unclear and may be connected with unknown underlying structures in the patient set.

High model accuracy in some patients, together with the error estimation procedure with model training and testing on separate patients, suggests similar tremor-correlated signal prop-

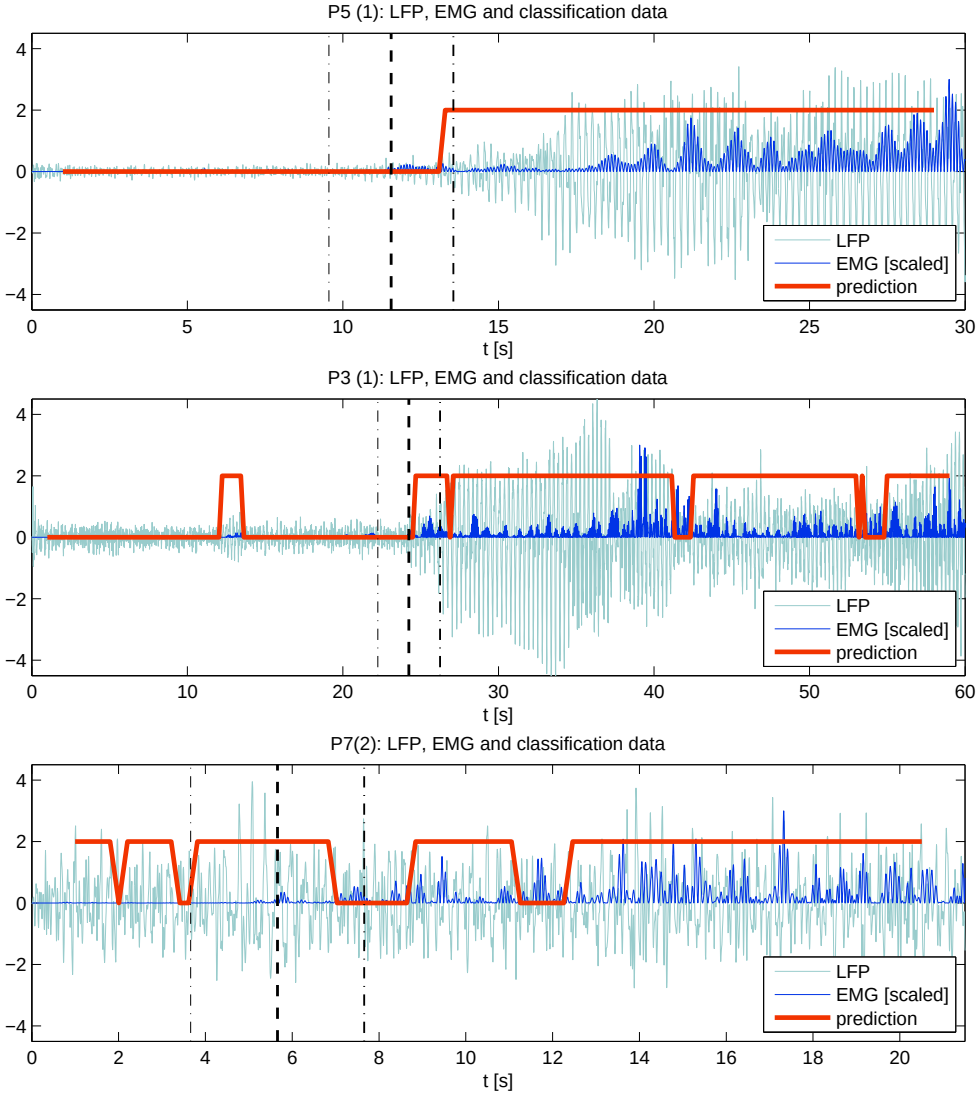

Figure 3.3: Tremor prediction on three patients, unseen previously by the classifier. The original LFP signal is shown along with flexor EMG activity and model output. EMG-based onset time and corresponding onset range are depicted by the vertical dashed lines.

erties across the whole dataset. On the other hand, the low system accuracy in some other patients suggests these tremor-bound properties are shared by only a subgroup of all patients. Unfortunately, the size of the dataset does not allow for drawing conclusions about different patient subgroups, based on the tremor-related properties of LFP signals.

Another effect that could have contributed to the mixed results of different patients might be the method of specification of tremor onset from EMG data. Possible patient variability in time shift between the outbreak of tremor-related changes in the LFP signals and tremor onset detected in the EMG activity may shift the classification accuracy significantly. This could be one explanation for the extremely poor specificity in patient P8: the results might be improved if the EMG tremor onset were marked earlier. To achieve accurate tremor detection, manually marked tremor onset, based on observation of the patient would be needed.

This also applies to classification accuracy based on LFP from different nuclei: even though the classification accuracy for both patients with GPi implants was relatively high, the lack of data does not allow for more general conclusions. The small amount of data was the reason why treating all available patients as one coherent group was the only option for unbiased error estimation and was therefore chosen in this study.

Unlike epileptic pre-seizure periods, Subthalamic nuclei in the Parkinsonian brain are thought to display tremor synchronization activity only seconds (rather than minutes) before severe muscle tremor starts (Brown 2003; Wang et al. 2007).This places greater responsibility and urgency on the classification algorithm. In this study, tremor period annotation was calculated from filtered normalized forearm EMG (described above). Even though the calculation was designed to be patient-independent, the precision of such a detection procedure can only be accurate to a certain extent. Possible variance of the delay between outbreak of tremor synchronization activity in the LFP signals and actual muscular tremor may be the source of improper annotation and thus increased classification error.

In the tremor detection task, sensitivity is of primary importance, as it represents the retention of the permanent beneficial effect of the stimulator. Specificity in this case represents battery saving capabilities for the system and is therefore not as important. Thus, an ideal classification system should be tuned for maximum possible sensitivity to retain an equivalent effect to that of full-time stimulation. Due to the intermittent nature of PD tremor in the observed patients, long-term LFP recordings would be necessary in order to obtain proper sensitivity and especially specificity estimates.

This study has focused strictly on analyzing and detecting tremor periods. The basis for this being a desire to understand more fully the links between tremor onset and LFP signals. This is a clear step towards accurately predicting tremor onset such that a signal can be employed as an adaptive trigger to fire the stimulator as part of a feedback loop before tremors actually occur. Studies are also ongoing into tremor prediction (Pan et al. 2012; Wu et al. 2010) the desire

here however is to provide a stronger base for further research along these lines. Despite the limitations caused by the relatively small amount of data, the accuracy estimation on unseen patients which was implemented in this study should provide unbiased results. Moreover, the results show that a universal classifier without the need of parameter tuning for specific patient is feasible with satisfactory results at least for a subgroup of patients. The requirements that have to be fulfilled for such a detector to be broadly applicable remain among the research questions yet to be answered.

3.4 Chapter conclusions

A system for online detection of Parkinson disease tremor, based on LFP signal features of different kinds, was proposed in this study. A global classifier without the necessity of parameter-tuning was developed and tested. Despite the fact that a patient non-specific model was trained, very good results were obtained for four out of eight patients, supporting the feasibility of the multi-feature NN approach in tremor detection for some patients at least. The performance of the system on unseen patients showed that at least a subgroup of the patients could benefit from the system, if and when it was implemented in a stimulator device.

To draw more general conclusions about performance — especially stability and robustness — of the system, a much wider dataset, including data from different positions and situations in different individuals will be necessary. Larger amount of data could also provide a basis for further evaluation of individual feature properties, possible patient types or differences in classification accuracy between different targets. This is the subject of ongoing research.

One final comment is that it may well be the case that patients with PD can be grouped in terms of the LFP activity witnessed in atremorous, tremorous and onset time periods. Certainly from our limited studies thus far — both reported on here and elsewhere (Burgess et al. 2010) — this would appear to be a natural conclusion.

In our more recent study (Camara et al. 2015) we investigated this possibility of different types of resting tremor in PD patients. The experiment showed that clustering of tremor chunks based on basic signal features (energy, difference, entropy...) divided patients into two very consistent groups. Even thought the dataset was comparably small to the study presented in this chapter (7 patients), the results have shown that applying clustering algorithm to patient data prior to tremor detection may significantly improve the detection accuracy and improve performance of a future closed-loop DBS system. A question underlying this conclusion however is whether or not the existence of such groups can be associated with underlying neurological relationships.

Chapter 4

Artifact identification in microelectrode recordings

Chapter summary

The DBS exploration procedure represents a unique possibility to record neural activity of human deep brain structures. Despite this unique potential the data possess, the main factors affecting the clinical procedure are the best possible surgery outcome and safety of the patient. The MER recordings may, therefore, contain a relatively large amount of external noise, caused by patient movement, speech or vibrations of the stereotactic frame and electromagnetic interference, which may have a harmful effect in subsequent data analysis.

This chapter presents a set of novel classifiers for identification of the MER artifacts, evaluates them on an extensive database of MER data and compares them to existing methods. Further, the chapter presents the *sigInspect*: a graphical user interface tool for MER inspection and manual or automatic artifact annotation. The main purpose of the tool lays in automatic or manual data preprocessing for scientific purposes, and to the best of our knowledge, this tool is the first available utility of its kind. Despite the main aim being the MER signals, the tool can be used for any time series, consisting of one or many parallel channels.

The research presented in this chapter is about to be submitted as:

- Bakstein et al. (2016) Automatic Artifact Detection in Microelectrode Recordings, in *Journal of Neuroscience Methods*

and uses also research presented in:

- E. Bakstein, J. Schneider, T. Sieger, D. Novak, J. Wild, and R. Jech (2015). "Supervised segmentation of microelectrode recording artifacts using power spectral density". In: *Proc. of 37th Annual International Conference of the IEEE Engineering in Medicine*

and Biology Society, At Milano, Italy. Vol. 2015-Novem. IEEE, pp. 1524–1527. DOI: 10.1109/EMBC.2015.7318661

The *sigInspect* tool is available at `https://github.com/ebakstein/sigInspect`

4.1 Introduction

Extracellular microelectrode recording (MER) using electrodes with a tip size around 1 μm (Slavin et al. 2004) is a basic technique for acquiring micro-EEG (μEEG) activity at the level of detail of individual neurons (*single-unit activity*). Due to the small size of the electrodes and low voltage of the source signal, MER recordings are susceptible to mechanical shifts and electromagnetic interference which result in signal artifacts (Stacey et al. 2013). While some components of the external noise can be filtered easily — e.g. 50 Hz or 60 Hz mains noise filtering using a notch or comb digital filter or sophisticated hardware design (Obien et al. 2015) — other may be more difficult to define and suppress.

In this study, we describe the aspects of the most prevalent artifacts, as observed on an extensive MER database, obtained during Deep Brain Stimulation (DBS) microrecording. As the DBS technique has been used routinely in therapy of neurodegenerative disorders for more than two decades (Benabid et al. 1996) and microrecording is still used in a vast majority of DBS centers worldwide (Abosch et al. 2013), the DBS surgery serves also as a prominent source of human sub-cortical μEEG data for scientific purposes. Despite the value the DBS MER signals may have for the research community, it is the clinical aspect which mostly determines the procedure and puts strain on available time and instrumentation. The sources of undesired artifacts in DBS surgery include electrical appliances in the operating theatre, electrode vibration (after manual electrode shift or touches to the microdrive or stereotactic frame) and movement or speech of the patient. Therefore, the artifacts in DBS MER data are common, which we also illustrate on data samples from four DBS centers, comprising 121 microexploration trajectories from 69 Parkinson's Disease patients.

The presence of artifacts in a MER signal may have a dramatic impact on subsequent signal processing, such as spike detection and sorting. Its severity will depend on the particular processing pipeline as well as on the character of the artifact. In the case of single or multi-unit analysis, a spike detection and spike sorting methods are used to separate activity of neurons close to the electrode tip from the background activity (i.e. the net activity of neurons further away from the electrode — see (Martinez et al. 2009; Mesa et al. 2013) for model, explaining signal generation). Commonly used extracellular spike detection methods use amplitude distribution to estimate appropriate value of the detection threshold (Harris et al. 2000; Quiroga

et al. 2004; Rutishauser et al. 2006) and are therefore sensitive to background noise level (Wild et al. 2012), as well as artifacts.

One possible approach to suppress the effect of exogenous noise is to use a robust estimator for background noise level (Dolan et al. 2009). However, this may lead to a large number of noise peaks exceeding the threshold, introducing false positive candidate spikes to the subsequent step of spike sorting. False positive candidate spikes can be partially resolved by discarding spikes of anomalous shape (Quiroga et al. 2004; Lourens et al. 2013) but still represent a significant source of undesirable noise in the whole analysis pipeline, which may lead to a loss of sensitivity or noise in the resulting spike trains.

4.1.1 Existing approaches

The prevailing approach to attain an artifact-free dataset in MER-based studies relies on manual inspection and disposal of contaminated signal segments (Zaidel et al. 2010; Seifried et al. 2012). However, this approach may be lengthy and still not provide optimal results — some artifacts can not be identified from the time series only due to their low projection to signal envelope and other modalities including spectrogram and audio playback are necessary to identify all artifacts. Therefore, many researchers use their own (semi) automatic methods, ranging from simple amplitude thresholding (Weegink et al. 2013) through statistical testing of amplitude distribution in short signal windows (Moran et al. 2006; Zaidel et al. 2009) to sophisticated amplitude and power spectral density (PSD) based systems (Moran et al. 2008; Cagnan et al. 2011; Verhagen et al. 2015). The threshold parameters are usually selected to match subjective evaluation of the investigator.

Two methods for identification of clean MER segments were previously published in detail with performance evaluation, both based on stationary segmentation, i.e. searching for segments consistent in a selected signal feature. The first of these methods is based on variance of the autocorrelation function (Falkenberg et al. 2003; Aboy et al. 2006), the second on variance of signal wavelet decomposition (Guarnizo et al. 2008). These methods, while suitable for rapid changes in amplitude (as presented on simulated data in the original publications), seem to be less appropriate for motion artifacts and electromagnetic interference (Bakstein et al. 2015) — two of the most prevalent noise types observed on our MER recordings database.

While artifact detection toolboxes are common for other broadly available signals such as the electro-encephalogram (Lawhern et al. 2013), to the best of our knowledge, the microelectrode-oriented toolboxes aimed so far at stimulation or common-noise artifacts only and are therefore not suitable for general-purpose MER preprocessing.

4.1.2 Proposed artifact detection method

As we showed in our recent paper (Bakstein et al. 2015), very good detection results compared to existing solutions can be achieved by a simple linear classifier based on power spectral density (PSD) of the signal. In this chapter, we extend this approach by presenting an artifact detection model, based on multiple time domain and spectral features. We tested a range of models based on decision trees, support-vector-machines (SVM) and boosting and evaluated their performance on a datatabase collected from four DBS centers. We make the resulting classifiers available as a part of *sigInspect*: a Matlab tool for semi-automatic signal classification. The *sigInspect* allows users to visualize single or multichannel MER data, generate initial data annotation using a selected algorithm and possibly correct the result manually.

As opposed to artifacts from concurrent electrical stimulation, which can be well described and are sufficiently studied in the literature (Egert et al. 2002; Wagenaar et al. 2002; Wagenaar et al. 2005; Obien et al. 2015), as well as artifact detection methods based on blind source separation or inter-electrode correlation, which can be applied to microelectrode arrays (Gligorijevic et al. 2009; Paralikar et al. 2009), the main scope of this chapterare the still heavily used single-channel MER data (i.e. one channel, or multiple electrodes spaced away in the order of mm or cm), such as those obtained during the DBS microexploration.

Despite the main scope of this chapter and the presented automatic methods being the DBS MER data, the *sigInspect* is a general signal viewing and annotation tool that can be used for analysis and annotation of other types of single and multi-channel signals and is not restricted to MER only. We provide the *sigInspect* toolbox for free use in the research community.

4.2 Methods

This section provides overview of the data collection and annotation process, including description of the most commonly observed characteristic artifact types. Further, the features and classifiers are described and the section also provides introduction to the *sigInspect* toolbox.

4.2.1 Data collection

All data used in this study has been collected during electrophysiological exploration for deep brain stimulation surgery in four DBS centers in the Czech Republic and the Slovak Republic. All patients were implanted either unilaterally or bilaterally, using one to five microelectrodes in a cruciform configuration (the "Ben-gun"), spaced 2 mm around the central electrode. The system used in each center, together with sampling frequency and maximum number of microelectrodes used can be found in Table 4.3. Median recording length was 10 s in all centers, signals

shorter than 5 s were discarded. All centers used the Leadpoint recording system (Medtronic, MN), with the recorded signals sampled at 24 kHz. The recorded signals were band-pass filtered between 500-5000 Hz upon recording.

In all centers, data recording was part of an unaltered standard therapeutical procedure and as such, no ethical committee consent was necessary.

4.2.2 Artifact annotation

After initial informal tests and consultations with an expert neurologist, we decided to base the annotation on visual and auditory inspection of signal time series: signal plot and audio playback of signal peaks above user-adjustable threshold, motivated by clinical experience with microrecording software used during the surgery (Medtronic LeadPoint). Additionally, the user was provided with a spectrogram heatmap, showing short-time Fourier transform spectra on parallel time scale with the time series.

All artifact annotations throughout this text refer to labeling of one second signal segments. Even though we have undertaken experiments with exact artifact start-end labeling, the process was laborious and the overlap between multiple annotators was low, as determination of the exact start or end time point turned out very difficult and unclear in many cases. In cases where the data included multiple channels from electrodes recorded simultaneously, all channels were visualized in parallel for easier identification of movement artifacts, often spanning across multiple channels.

Annotator team synchronization

In the initial phase, a set of 100 ten-second single channel MER signals has been annotated by the expert. The expert annotation has been reviewed and discussed among the whole team of eight raters and used to describe characteristic manifestation of the most common artifact types. The resulting consensus was then converted to an artifact catalog document, shared among the team. In subsequent phase, all team members annotated the same set of 20 multi-channel signals in order to ensure agreement on artifact and clean signal definition. The resulting annotations have been compared against consensus annotation (obtained by majority vote of all members' annotations) and discussed. The process was repeated twice to achieve better agreement.

Once the team has achieved agreement, all signals available for annotation were distributed among the team. As the initial database contained more than 1600 multi-channel recordings, only a small subset of about 2% of signals has been shared among the team as a proofing sample. Evaluation of team annotation on this proofing set revealed two out of eight raters with very low sensitivity and one member having poor accuracy overall. Based on these results, the three members were excluded from the annotation team and the rest of MER recordings has been

divided among the remaining five members. Additionally, all signals assigned to the poorly performing raters have been re-annotated by the final team. At last, the five team members evaluated signals from the remaining centers; dataset for each center contained approximately 5% common signals as a separate proofing sample.

Observed artifacts

For further analysis and evaluation, the observed artifacts can be grouped into the following clusters (cluster shortcut in parenthesis)

- Mechanical artifact, manifested usually by short-time, high-power signal peaks, usually spread across the frequency spectrum. (POW)
- Low-frequency interference below the mains frequency (50 Hz), causing visible variation in signal offset or baseline ($BASE$)
- Electromagnetic interference at one or multiple stable frequencies, well localized in a narrow band in the frequency spectrum and stable over time ($FREQ$). Frequency of the observed long-term interference often differed from the expected odd harmonics of the mains frequency (50 Hz, 150 Hz, 250 Hz etc)
- irritated neuron: spiking activity of very high and variable amplitude and firing rate ($IRIT$)
- other artifacts that cannot be assigned into any of the groups above. ($OTHR$)

The MER signal may contain one, as well as several artifact types at the same time. Clean signal (CLN) is defined as the absence of artifacts.

4.2.3 Automatic classification methods

This section provides overview of all artifact classifiers and methodology used for parameter optimization and classification. Due to relatively low agreement on exact artifact type between different annotators (see Section 4.4.2 for details), all classifiers were designed only as two-class classifiers, trained to distinguish clean signals (CLN) from signals with all other artifact types.

Stationary segmentation methods

Two stationary segmentation methods, based on division of signals into short segments, were described previously by i) Falkenberg and Aboy et. al. (Falkenberg et al. 2003; Aboy et al. 2006) based on variance of signal autocorrelation function (denoted COV) and ii) by Guarnizo et. al. (Guarnizo et al. 2008) based on variance of signal wavelet decomposition (denoted

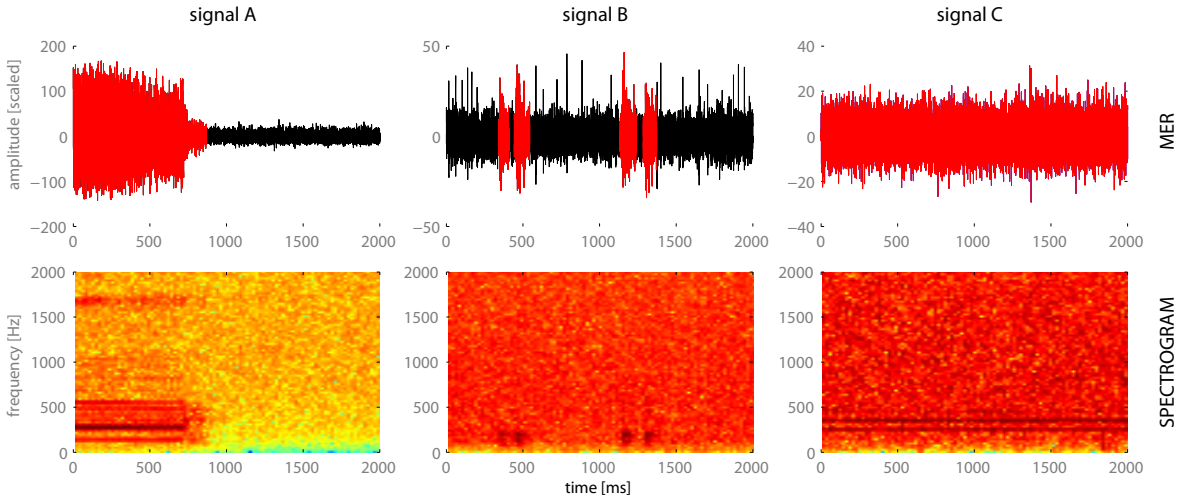

Figure 4.1: Two second examples of most commonly observed artifacts: raw MER signal with artifact regions in red (top row) and corresponding spectrogram (bottom row). Signal A) represents intermittent electromagnetic interference, signal B) mechanical artifact and signal C) uninterrupted electromagnetic interference at ca 235 and 350 Hz.

SWT). These methods compute variance ratio of neighboring segments and compare the value to a manually preset threshold. Points where the threshold is exceeded are marked as change points. In the end, the longest sequence of signal segments, uninterrupted by change points, is returned. An extension of these techniques has been presented in the paper (Bakstein et al. 2015) and is described in the next section.

Extension of the stationary segmentation techniques

In order to compare performance of the stationary segmentation techniques to manual annotation and other classifiers, we present extension of these methods from (Falkenberg et al. 2003; Aboy et al. 2006; Guarnizo et al. 2008). In their basic version, these methods first divide the signal $\boldsymbol{X}$ into m non-overlapping segments $X_1, X_2 \ldots X_m$ and compute statistics $\gamma(X_i)$ for each segment, where $\gamma(\cdot)$ is autocorrelation function of the segment (COV) or stationary wavelet transform (SWT). In the next step, variance of each transformed segment is calculated according to

$$v_i = var\{\gamma(X_i)\}), i \in \langle 1, m \rangle, \tag{4.1}$$

Variances of neighboring segments are then compared according to:

$$d_{i,j} = \frac{max(v_i, v_j)}{min(v_i, v_j)}, i \in \langle 1, m-1 \rangle, j = i+1 \tag{4.2}$$

Divisions between segments with distance statistic d_{ij} exceeding a manually pre-chosen threshold Θ then determine breakpoints between stationary segments. The longest stationary segment is found and returned.

We further extend this method by computing distance between all possible segment pairs, forming a distance matrix

$$D = \begin{pmatrix} 0 & d_{1,2} & \cdots & d_{1,m} \\ d_{2,1} & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ d_{m,1} & d_{m,2} & \cdots & 0 \end{pmatrix}. \tag{4.3}$$

Note that due to properties of the distance measure from Eq. 4.2 the matrix is symmetric with $d_{ii} = d_{ji}$. In the next step, all values d_{ij} exceeding the classification threshold Θ are replaced with zeros, other with one, leading to a graph, defined by the following adjacency matrix:

$$E = \begin{pmatrix} 0 & e_{1,2} & \cdots & e_{1,m} \\ e_{2,1} & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ e_{m,1} & e_{m,2} & \cdots & 0 \end{pmatrix}, e_{i,j} = \begin{cases} 1, & \text{if } d_{i,j} < \Theta \\ 0, & \text{otherwise} \end{cases} \tag{4.4}$$

The graph, represented by adjacency matrix E is then scanned for maximum component. With this modification, the algorithm returns the largest component of similarity in the original signal, which may be even a non-contiguous signal subset. The procedure used to search the maximum signal component from the adjacency matrix is outlined in the Matlab-style pseudocode in Algorighm 1. Note that the method only requires all segments to be connected by a non-interrupted path, sub-threshold similarity between all possible segment pairs in the component is not required. Therefore, value of the optimal detection threshold will also differ from the originally published methods.

This method should provide results much closer to manual signal annotation, especially in the case of short-term artifacts. Also in analyses where signal contiguity is not required (such as in background activity feature calculation), this approach may minimize the amount of unnecessarily removed data.

In this chapter, we optimized three parameters of each algorithm: i) segment length (0.25, 0.33, 0.5 or 1 s) ii) detection threshold and iii) the number of segments within one second window labeled by the classifier as artifact, necessary to mark the whole second as artifact — the last point was necessary since the manual annotation labels were available for one-second windows only.

Algorithm 1 Identification of the maximum component from the adjacency matrix

input: E; m*m adjacency matrix, m is number of segments
output: *max_comp*; indices of maximum component

comp =**zeros**$(1, m)$; % denotes which segment belongs to which component
act_comp = 1; % Actual component
while any(*comp* == 0) **do**
  % loop as long as there are unassigned segments
  open = first zero in *comp*
  closed = [ ]
  while NOT isempty(open) **do**
    comp(*open*(1)) = *act_comp* % assign actual comp. to actual segment
   % Expand current state (all segments adjacent to current segment)
    children = **find**(E(*open*(1),:))
   % take the first element from open, find to which segments exists a direct path
    for *ch* in *children* **do**
      if *ch* not in *open* **OR** *closed* **then**
        open = [*open ch*] % add ch to open
      end if
    end for
   % Move current node from open to closed
    closed = [*closed cur*]
    open = *open*(2:end)
  end while
  act_comp = *act_comp* + 1
end while
  % Find the largest component
comp_len = **zeros**(1,*act_comp*)
for *cur* in 1:*act_comp* **do**
  comp_len(*cur*) = **sum**(*comp* == *cur*)
end for
[∼,*max_comp*] = max(*comp_len*)
return *max_comp*

Maximum spectral difference method

A simple detection method, based on the power spectral density of MER signal, has also been presented in our aforementioned paper (Bakstein et al. 2015). The basic assumptions are that a PSD of a clean band-pass filtered MER signal is smooth, contrary to most signals with artifacts, which commonly contain high peaks and other disturbances. In the first step, a mean spectrum *clSpec* of clean signal segments is calculated from a set of n training signals $\boldsymbol{X} = \{X_1, X_2, ..., X_n\}$ with corresponding artifact annotation $\boldsymbol{a} = \{a_1, a_2, ..., a_n\}$ with a_i equal to 1 for clean signals and 0 for signals with artifacts, according to:

$$clSpec = \frac{1}{\sum_{i=1}^{n} a_i} \cdot \sum_{j=1}^{n} a_j \, NPSD_M(X_j), \tag{4.5}$$

where $NPSD_M(X_j)$ is the normalized power spectral density of signal segment X_j, computed using Welch's method with discrete Fourier transform of length M and normalized according to:

$$NPSD_M(X_j) = \frac{PSD_M(X_j)}{\sum_1^M PSD_M(X_j)} \tag{4.6}$$

where $\sum_1^M PSD_M(X_j)$ is sum of all psd spectrum bins — such that the sum of the resulting spectrum is equal to one and is therefore independent of total signal power. In our experiments, the length of discrete Fourier spectrum was set to $M = 2048$, as well as the window length, window overlap was set to 50%, sampling frequency in all centers 24 kHz.

To compare the sample PSD spectrum to an unseen signal segment Y, maximum absolute difference from the sample spectrum has to be calculated according to:

$$d = \max |\, NPSD_M(Y) - clSpec \,| \tag{4.7}$$

Optimal detection threshold for d can then be determined on a training dataset and used for classification. Clean $NPSD$ spectrum, as well as spectrum for the three major artifact types can be found in Fig. 4.2.

Multi-feature classifiers

In addition to the simple detection methods mentioned above, we have implemented a range of classification methods, based on multiple features, derived from raw MER signal and its normalized power spectrum. The features were designed in order to describe the most prominent properties of various artifact types, compared to clean MER signal. Characteristic MER spectra of different artifact types can be found in Fig. 4.2. The normalized power spectral density was first computed according to Eq. 4.6 using Welch's method with Fast Fourier Transform of length 2048 (equal to window length) and 50% overlap. The parameter choice was based on preliminary tests and observations on the Prague-CV database. The first set included 17 features and was

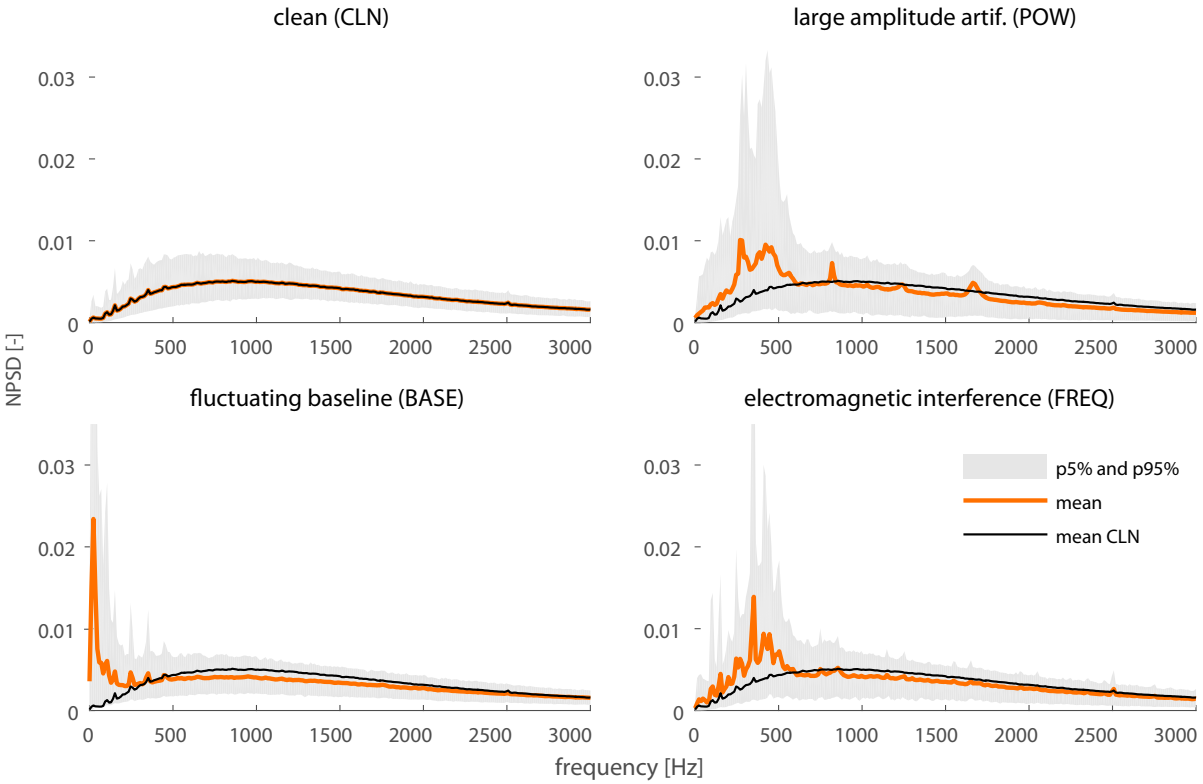

Figure 4.2: Normalized power spectral density for different artifact types, each computed from 1000 randomly selected signals from the Prague-CV dataset. Mean value for each artifact type (orange) with 5% and 95% percentile for each spectral bin is compared to the $NPSD$ of clean signals (black).

extended to the total of 19 features after evaluation of initial tests on the Prague-CV database. The features were calculated from one second signal segments to match temporal resolution of the annotation. All features in the feature set are summarized in Table 4.2.

As seen from the description of individual features, high correlation is to be expected in many cases, such as the *psdMaxStep* and *psdMax* (due to very sharp character of spectral peaks in signals with electromagnetic interference), or *sigP90* and *sigP95* (due to smooth character of signal amplitude distribution in lower percentiles). Therefore, all selected classification methods have to perform some sort of feature selection, allowing for correlated features. The multi-feature classifiers implemented include:

- **Decision tree** classifier with limits on minimum parent node and leaf size and different splitting criteria.

- **Support Vector Machine** (SVM) classifier with linear or radial-basis kernel using different optimization methods and kernel properties. The SVM classifier was preceded by feature selection step — see description below.

- **Boosting** classifiers, using different algorithms (AdaBoostM1, LogitBoost, GentleBoost, RobustBoost, Bagging) and varying learning rate. The weak learner used was a decision tree.

All classifier parameters and ranges in which they were optimized can be found in Table 4.1.

Table 4.1: Overview of optimized classifier parameters and values

classifier	optimized parameters
stationary segmentation	*method:* i) COV (covariance, Aboy) or ii) SWT (wavelets, Guarnizo) *segment length:* {.25,.33,.5,1} *aggregation threshold:* for .25 s window: {1, 2, 3, 4}, for .33 s {1, 2, 3}, for .5 s {1,2} *threshold for COV:* <.8, 3.5>in .1 steps *threshold for SWT:* <9.5, 13>in .1 steps
diffPSD	*threshold:* <0,0.025>in 0.0005 steps
decision tree	*split criterion:* i) Gini's diversity index ii) max. deviance reduction *min size of parent node:* {1, 100, 200, ...,500, 100, 1500, 5000} *min. leaf size:* {1, 250, 500, 750,...,2500} maximum up to half of current parent node min. size
SVM	*feature selection criterion:* quadratic, linear, mahalanobis *feature selection stopping tolerance:* {.001, .005, .01} *SVM method:* i) Sequential minimal optimization or ii) least squares *SVM kernel:* i) linear ii) radial basis function (RBF) *SVM kernel sigma (only for RBF):* {.5,1,2}
Boosting	*algorithm:* AdaBoostM1, LogitBoost, GentleBoost, RobustBoost, Bag *number of learners:* {10, 20,...,50,75,150,200,250} *learning rate:* {.1, .4, .7, 1} *RobustBoost error goal:* {.05, .1, .15, .2} *RobustBoost error max margin:* {.01, .05,.1}

The SVM classifier was preceded by a feature selection step in order to reduce the number of features, as well as their redundancy. We used forward wrapper feature selection with discriminant analysis classifier. The algorithm starts with an empty feature set and adds a single feature that provides best accuracy using the selected classifier. Then, all possible sets of two features including the already selected feature set and a candidate feature from the remaining part of the features. The best-performing two-feature set is fixed and the process continues

with three-feature sets and on, until stopping criterion (minimum improvement in classification accuracy) is achieved. The classifier used for feature selection was discriminant analysis (linear, quadratic or mahalanobis distance-based) and we used internal 5-fold crossvalidation to estimate out-of-sample performance of the classifier. The classifier type, as well as value of the stopping criterion was subject to parameter optimization, as seen in the Table 4.1.

4.2.4 Crossvalidation scheme

In order to evaluate classifier performance on unseen data, we divided the training data from Prague database into two datasets: data from three DBS patients (6 multi-electrode trajectories) were kept aside for final classifier testing as the Prague-TS set, while the remainder, denoted Prague-CV, was used for feature evaluation, classifier training and parameter optimization. The test set was then extended with data from three additional DBS centers to evaluate classifier performance in a real-life scenario; mimicking an independent research group adopting the provided classifiers on their own data.

The main crossvalidation procedure (denoted A) was following:

1. **Ten-fold crossvalidation**: The Prague-CV dataset was divided randomly into 10 subsets. In each iteration, all parameter combinations were trained on 9 subsets and validated on the remaining one. Confusion matrix on the validation sample has been stored and the algorithm continued with the next iteration.

2. **Parameter optimization**: For each classifier type, the parameter set which optimized the validation performance was selected.

3. **Final classifier training**: Each classifier was trained on the whole Prague-CV dataset, using optimal parameters, obtained in the previous step.

4. **Out-of-sample performance estimation**: The final classifiers were used to classify data from the four test datasets and final performance evaluation was stored.

Overview of the crossvalidation procedure is also available in Figure 4.3. The presented crossvalidation scheme was chosen in order to provide as unbiased estimate of classification performance on unseen data (i.e. data from unseen subjects, as well as unseen centers) as much as possible.

As the dataset contained a notable class imbalance (about 75% of clean signals; see section 4.4.2 for details), we considered the classification error to be an inappropriate performance measure to be optimized (75% accuracy achievable by labeling all signals as clean). Therefore, we chose the Youden's J-statistic, computed as

$$J = sensitivity + specificity - 1. \tag{4.8}$$

Table 4.2: Feature set overview

feature	definition	rationale
pow	signal power	higher overall power in signal windows with artifacts
powDiff	maximum power difference between adjacent 0.05s signal segments	power artifacts abrupt in time
sigP90, sigP95, sigP99	raw signal amplitude percentile (90th, 95th,99th)	artifacts commonly include very large peaks
ksnorm	value of the KS statistic of a one-sample Kolmogorov-Smirnov normality test	pronounced non-normality in signals with artifacts
maxCorr	maximum correlation coefficient among multiple signal channels in 0.05 signal segments	mechanical artifacts often spread across channels and cause high outlier values and thus high correlation
psdP75, psdP90, psdP95, psdP99	percentile of NPSD	artifacts localized in spectrum - high spectral peaks
psdMax, psdStd	maximum and standard deviation of NPSD	global PSD description
psdMaxStep	maximum difference between adjacent bins of the NPSD	artifacts localized in spectrum - sharp spectral peaks
psdF100	maximum of the NPSD below 100 Hz	baseline artifacts at a well localized frequency
psdFreq	maximum of the NPSD, divided by median value below 5 kHz	additional normalization of the NPSD
psdPow	maximum of the NPSD in range 60-600 Hz, divided by mean NPSD between 1 and 3 kHz	power artifacts very common in this range
psdBase	maximum of the NPSD in range 1-60Hz, divided by mean NPSD between 1 and 3 kHz	baseline artifacts, normalized
maxAbsDiffPSD	maximum absolute distance between NPSD and mean NPSD of clean sample signal segments	high artifact peaks in PSD

The convention used in this chapter denotes artifact-contaminated signal samples as a positive class. Thus, the sensitivity, also called the true positive rate, is computed as the ratio of correctly classified artifact samples (true-positives, TP), divided by the count of all artifact samples ($sens. = TP/P = TP/(TP + FN)$), where FN is the false negative count, i.e. the number of artifact samples, incorrectly classified as clean. Specificity is then the false negative rate, computed as the ratio of correctly classified negative examples (clean signals, true negatives, TN) to all negative examples ($spec. = TN/N = TN/(TN + FP)$), where FP are false positives (clean signals incorrectly marked as artifacts).

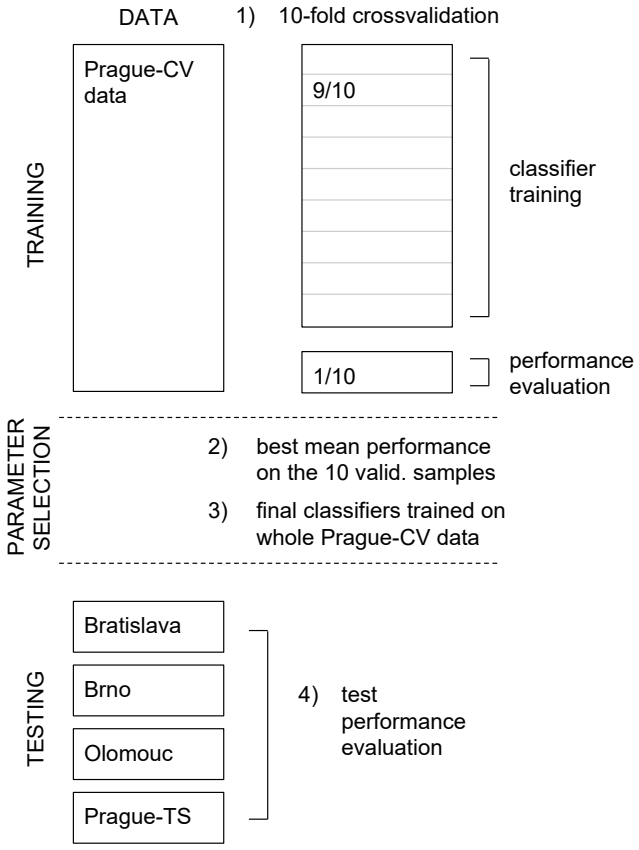

Figure 4.3: Overview of the main crossvalidation scheme A: 1) All parameter combinations are trained in a ten-fold crossvalidation and evaluated on ten validation samples. 2) Final parameter combination for each classifier type is selected to maximize overall performance on the 10 validation samples. 3) Classifiers are re-trained using all Prague-CV data. 4) Out-of-sample (test) error is estimated using the four validation datasets.

Additional post-hoc crossvalidation B.

Although our database contained data from four centers, the main crossvalidation procedure utilized only data from a single center for classifier training. We have therefore implemented an additional crossvalidation post-hoc test to identify the impact of addition of data from multiple centers into the training set. In this leave-one-subject-out (LOSO) scenario, the classifiers were trained on all but one patients from the test set (Bratislava, Olomouc, Brno and Prague-TS) and validated on the remaining data from a single patient. The process was repeated until all patients were used for validation. Afterwards, the validation performance from all folds was aggregated.

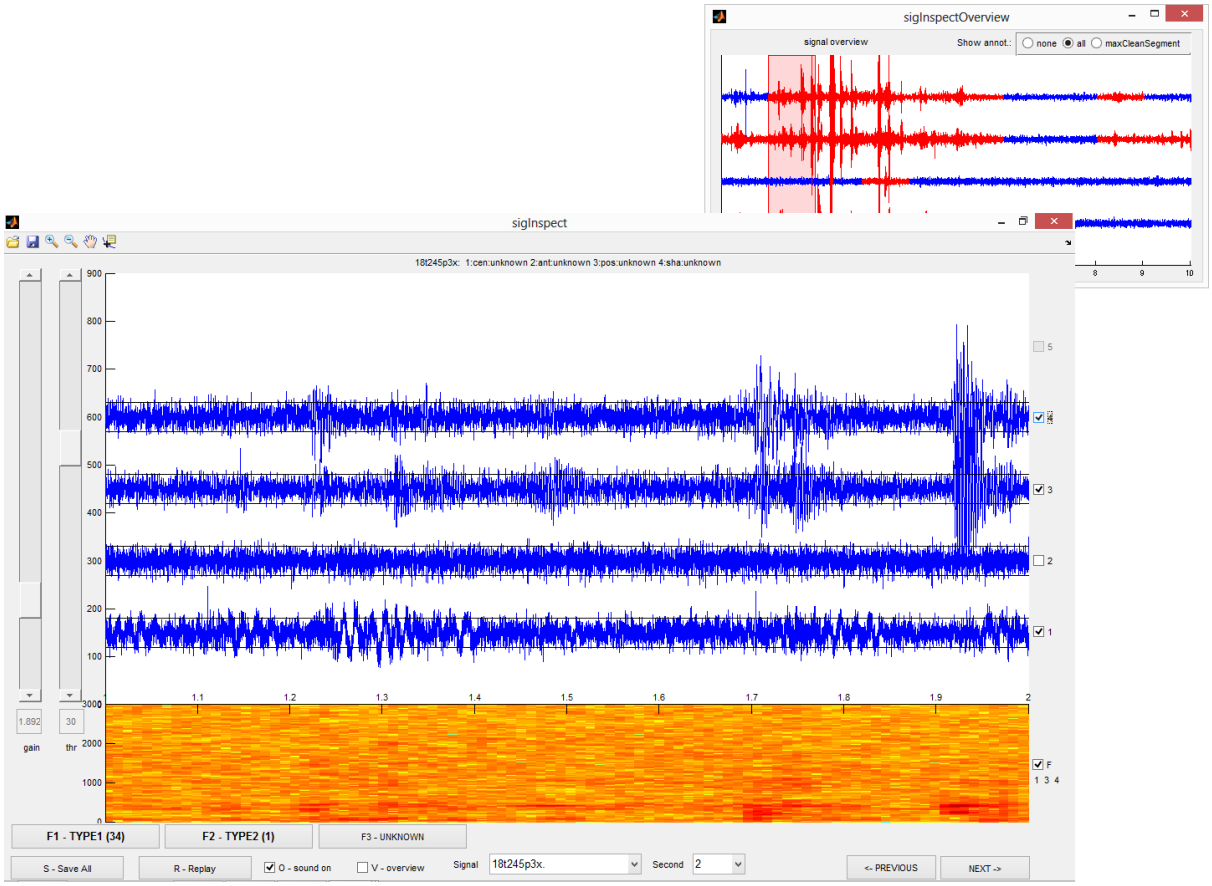

Figure 4.4: Main window of the sigInspect GUI, showing one-second sections in four parallel MER channels together with a spectrogram of selected channels. Artifacts can be seen in channels 1, 3 and 4, while channel 2 represents normal activity. A secondary window, showing overview of the whole signal with second annotated as artifacts in red can be seen in the upper right corner.

4.3 SigInspect: a Matlab GUI for MER inspection

The *sigInspect* is a Matlab GUI tool for viewing and analysis of multi-channel signals with particular focus on extracellular microelectrode recordings. The basic functionality allows visualization, playback and annotation of signal segments containing artifacts or other events. Advanced features include automatic annotation of MER artifacts, easy integration with custom databases and data formats through the `sigInspectDataInterface` and high level of configurability, including default configurations for different data interfaces.

This section provides a very brief overview of the sigInspect functionality. Full description and documentation can be found at the sigInspect page: `https://github.com/ebakstein/sigInspect`. The software is available for free download and use under the LGPL license.

4.3.1 Loading and viewing a signal

The simplest way to start viewing and annotation is to call `sigInspect` with signals as a parameter. Multiple signals can be passed in a cell array instead - user can then choose the displayed signal using the "signal" selector. The second parameter is sampling frequency in Hz (default value: $24kHz$ is set if second parameter is omitted)

Code 4.1: Basic data loading in the sigInspect

```
% 1 - single multi-channel signal (matrix as input)
size(signal)    % C x N matrix (C = channels, N = samples )
sigInspect(signal, samplingFreq);  % signal: channels in rows
                                   %         samples in columns
% 2 - multiple signals (cell array as input)
s={signal1,signal2,signal3};
sigInspect(s, samplingFreq);
```

After initialization, the main window appears as in Figure 4.4, showing the first one-second segment of the multi-channel signal. The user can than adjust signal amplitude and threshold for audio playback using GUI sliders or keyboard shortcuts. A set of checkboxes for channel selection, corresponding to each channel are placed at the right side of the main window. Alternatively, the user can use the number keys 1 up to the number of channels to toggle selection of individual channels, or use additional shortcuts "A" for *select all* or "I" for *invert selection*. A spectrogram of the mixture of all currently selected channels can be seen at the bottom. Skipping between adjacent signal segments is done using the arrow keys or GUI buttons at the bottom right.

The signal annotation is done by either clicking on one of the artifact label buttons at the bottom of the main window or by pressing corresponding keys (F1 to F12, depending on artifact types specified in the setup). Annotation is assigned to all currently selected channels for current second. Any combination of artifact types can be assigned to each second of each channel.

Many properties of the *sigInspect* can be set up during initialization, such as the number and names of artifact types available for annotation, sampling frequency, maximum number of channels, availability of spectrogram view, automatic gain and other. Parameter setting can be done using additional parameters in sigInspect initialization, as well as using the `sigInspectDataInterface`, described below.

4.3.2 Advanced functionality

For the sake of easier setup and integration with custom data formats, the users can implement a custom data interface by inheriting the `sigInspectDataInterface` abstract class. The implementation has to include two methods: `getSignalIds()`, which returns a cell array of string identifiers of available signals and `getSignalsById(signalId)`, which returns the multi-channel signal, corresponding to given identifier. The source of data and retrieval method is up to the user and can range from workspace variables to a remote database connection. An example of data interface which loads a list of all `*.csv` data from a directory and performs on-demand data loading can be found in Code 4.2.

Code 4.2: Example sigInspectDataInterface for loading CSV files

```
classdef sigInspectDataCsv < sigInspectDataInterface
    % define class, inherit sigInspectDataInterface abstract class
    properties
        dirPath='';
    end
    methods
        % constructor - just store the path, set settings
        function obj=sigInspectDataCsv(dirPath)
            obj.dirPath = dirPath;
            obj.settings.SAMPLING_FREQ=6000; % 6kHz sampling rate
            obj.settings.PLOT_STEP=1.5;      % distance between channels on ...
                the y-axis
            obj.settings.ARTIFACT_TYPES={'Type A','Type B','Unsure'};
        end
        % return list of signal ids - load all csv files from a directory,
        % use filenames as signalIds
        function signalIds = getSignalIds(obj)
            lst=dir([obj.dirPath '/*.csv']);
            signalIds = {lst(:).name}';
        end

        % read signals based on signalId (=filename)
        function [signals chInfo]= getSignalsById(obj,signalId)
            chInfo='';
            signals=csvread([obj.dirPath '/' signalId]);
        end
    end
end
```

Once the interface is implemented, it can be used very easily according to Code 4.3. Note that setting of sigInspect parameters is possible in the data interface implementation, as well

as during initialization.

Code 4.3: Example of sigInspectDataInterface usage

```
% initialize interface using its constructor
intf=sigInspectDataCsv('csvDemo/');
% change additional settings (optional)
intf.settings.NORMALIZE_SIGNAL_PER_CHANNEL = 0;
% run sigInspect
sigInspect(intf)
```

The *sigInspect* distribution includes basic implementations such as the `sigInspectDataBasic`, used by the default constructor, `sigInspectDataMatDir`, used to load all mat files on a specified path or the `sigInspectDataCsv` shown above.

4.3.3 Automatic data annotation

The sigInspect includes several methods for automatic artifact annotation, based on the results presented in this chapter. In order to annotate a set of signals, the user uses an initialized instance of the `sigInspectDataInterface` as a parameter to the function sigInspectAutoLabel. The method than runs the selected detection algorithm on all signals provided by the interface and produces a file with annotation, which can be then loaded and visualized in the GUI, according to Code 4.4.

Code 4.4: Automatic data labeling

```
% use the Basic interface + data in mat-file
intf = sigInspectDataBasic('mySignals.mat');
% change the settings
intf.settings.ARTIFACT_TYPES={'Automatic','MyArtif1','MyArtif2'};
% run autoLabel with the default method
sigInspectAutoLabel(intf,'myAutoAnnot.mat');
% run sigInspect
sigInspect(intf);
% now click the "load" icon in the toolbar and open myAutoAnnot.mat
```

Full description of sigInspect use, as well as list of all available setup parameters can be found in the sigInspect manual, available at the sigInspect webpage mentioned above.

4.4 Experimental results and discussion

4.4.1 Overview of collected data

In total, we collected data from four clinical DBS centers, which were divided into five datasets. Typical (as well as median) recording length was 10 seconds with a small proportion of shorter recordings. Recordings below 5 seconds were excluded from the dataset. Whole multi-channel recordings were used during annotation (for easier annotation of artifacts affecting multiple channels), as well as input to feature extraction and classification. The set contained a total of 2236 multi-channel recordings (7561 single-channel signals) from 121 MER exploration trajectories of 69 patients. Total recording length was 74273 signal seconds, which is 20h 37min. Full information about all datasets can be found in Table 4.3. All centers included in the study used the Leadpoint (Medtronic, MN) micro-recording system with a set of one to five tungsten microelectrodes in a "Ben-gun" configuration, sampled at 24 kHz.

Table 4.3: Overview of collected data

dataset	No. pat.	No. traj.	No. pos.	No. signals	total length [s]	mean length [s]	max. chan.	sampling freq. [Hz]	System
Bratislava	3	6	278	716	6372	8.9	3	24000	Leadpoint
Brno	5	10	193	772	7707	10.0	4	24000	Leadpoint
Olomouc	3	6	89	338	3380	10.0	4	24000	Leadpoint
Prague TEST	3	6	144	720	7130	9.9	5	24000	Leadpoint
Prague CV	55	93	1532	5015	49684	9.9	5	24000	Leadpoint
TOTAL	69	121	2236	7561	74273	9.9	5	-	-

4.4.2 Data annotation evaluation

After the extensive artifact characterization and rater team synchronization procedure described above, all MER data has been annotated by a team of five well-performing raters. As a small 3-5% random sample of data from each center was common to all raters (the *proofing sample*), it was possible to evaluate the annotation accuracy achieved on different datasets. Using majority voting, a consensus annotation was created for each proofing sample and compared to individual annotations by all team members.

Overall accuracy of raters' match with the consensus annotation can be found in Table 4.4 together with Fleiss' kappa and overview of proofing samples in all datasets. A level of agreement with the majority voting annotation above 90% was achieved in the cases of three out of four databases, while the raters achieved a slightly worse concensus on the Olomouc

database with 87.4%.

Contrary to determination of clean signal/artifact on the initial database, which achieved a respectable 93.5% on the initial Prague proofing sample, the agreement on exact artifact type was evaluated as poor, mainly due to the presence of many borderline cases difficult to assign to a particular category. Therefore, the classifiers were trained to distinguish between two classes (clean/contaminated signals) and the presented proportion of assigned artifact type on each database is provided for informative purpose only.

The evaluation of artifact content in each database as seen in Figure 4.5 and Table 4.5 reveals a very high amount of artifacts in all databases: the percentage of clean signal seconds varied between 51% (Olomouc) and 74% (Prague-CV), which stresses the importance of data cleaning prior to further analyses. Artifacts of the *FREQ* type, representing electromagnetic interference of higher frequencies (hundreds of Hz) were the most prevalent type across all datasets, appearing in up to 39.6% of all signals on the Olomouc database. Second most prevalent artifact type was *POW*, which was found in up to 17.3% of signals in case of the Prague-TS database.

Table 4.4: Inter-rater agreement on the proofing samples (randomly selected data portions shared by all raters)

	No. sig. seconds	No. positions	artif %	aggrement %	Fleiss' kappa
Bratislava	391	15	42.7	94.1	.792
Brno	400	10	27.0	90.5	.581
Olomouc	180	5	45.0	87.4	.562
Prague	950	23	26.8	93.5	.722

Table 4.5: Percentage of assigned artifact type combinations in each dataset, CLN represents clean signal seconds.

	CLN	POW	BASE	POW BASE	FREQ	FREQ POW	FREQ BASE	FREQ POW BASE	IRIT	OTHR
Bratislava	55.9	4.1	0.7	0.2	23.8	13.0	0.2	0.9	0.7	0.4
Brno	67.6	2.3	8.8	1.0	11.7	2.9	3.2	0.8	1.1	0.4
Olomouc	51.0	7.9	1.1	0.2	27.0	10.6	1.2	0.8	0.1	0.1
Prague-CV	74.6	2.3	4.9	0.3	13.2	2.0	2.2	0.2	0.3	0.2
Prague-TS	60.7	6.6	3.0	0.5	17.3	9.6	1.3	0.6	0.3	0.1

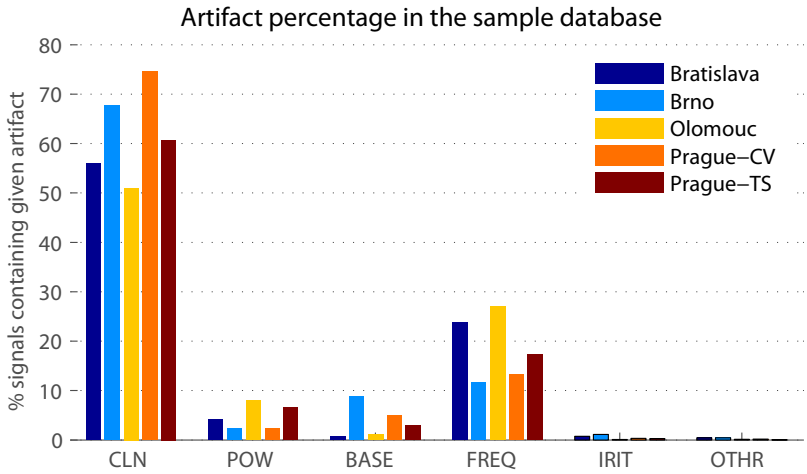

Figure 4.5: Percentage of artifacts of different types vs. clean signal (CLN) on a sample database of one second signal segments. As more than one artifact type may be assigned to each second, the sum of all percentages may exceed 100%. The exact percentage values can be found in Table 4.5.

4.4.3 Feature evaluation

The resulting feature set, based on temporal, spectral and statistical properties of the MER signal, consisting of 19 features, was a result of a multi-stage process. The initial feature set, containing 17 features was used for artifact classification on a sample database (a subset of the Prague-CV dataset) using decision tree classifier. The classification results were analyzed thoroughly, especially with respect to misclassified samples and classification accuracy across various artifact types. The results revealed false negative classification in cases of short-term events spanning multiple channels, which were similar in both temporal and spectral properties to physiological spikes, and therefore hard to detect. The feature set was thus extended with the *maxCorr* feature, based on multi-channel cross-correlation in short 0.05 s signal segments: the large-amplitude events synchronized in time in an otherwise stochastic signal cause outliers of a large value and high statistical leverage in comparison of corresponding parallel channel values. The correlation matrix among all channels then contains large maxima that can be used for detection of such events. Another additional feature was the *ksnorm*: the value of KS statistic of a one-sample Kolmogorov-Smirnov normality test. This feature was designed to distinguish abnormalities in the sample distribution of signal values, which is close to normal (or T-distribution due to heavier tails) in a clean signal.

Description and summary of the final feature set is given in Table 4.2, while evaluation of area under the receiver-operator characteristic (AUC) of each feature and histograms on the Prague-CV data can be found in Figure 4.6. As can be noted, the features based on steps and differences in PSD (*maxAbsDiffPSD, psdMaxStep, psdStd* and *psdMax*) showed best discrim-

inative properties with AUC values reaching up to 0.92, which may be considered very good. On the other hand, newly added features *maxCorr* and *ksnorm* showed relatively low detection capability with AUC values around 0.55 and 0.58 respectively. However, it has to be noted, that the AUC values reflect detection capability in a single-feature classification scenario and are therefore biased towards features designed for detection of the most prevalent artifact types — such as *FREQ* in the case of the spectral features. Also, the histograms presented suggest (and subsequent correlation analysis confirms) that some of the features are strongly correlated and overlapping. It can be assumed, that an appropriate feature selection method should be capable to select a feature combination consisting of complementary features, including features with high additional benefit to the feature set despite their low overall performance on the whole imbalanced dataset. Therefore, all 19 features were kept in the dataset.

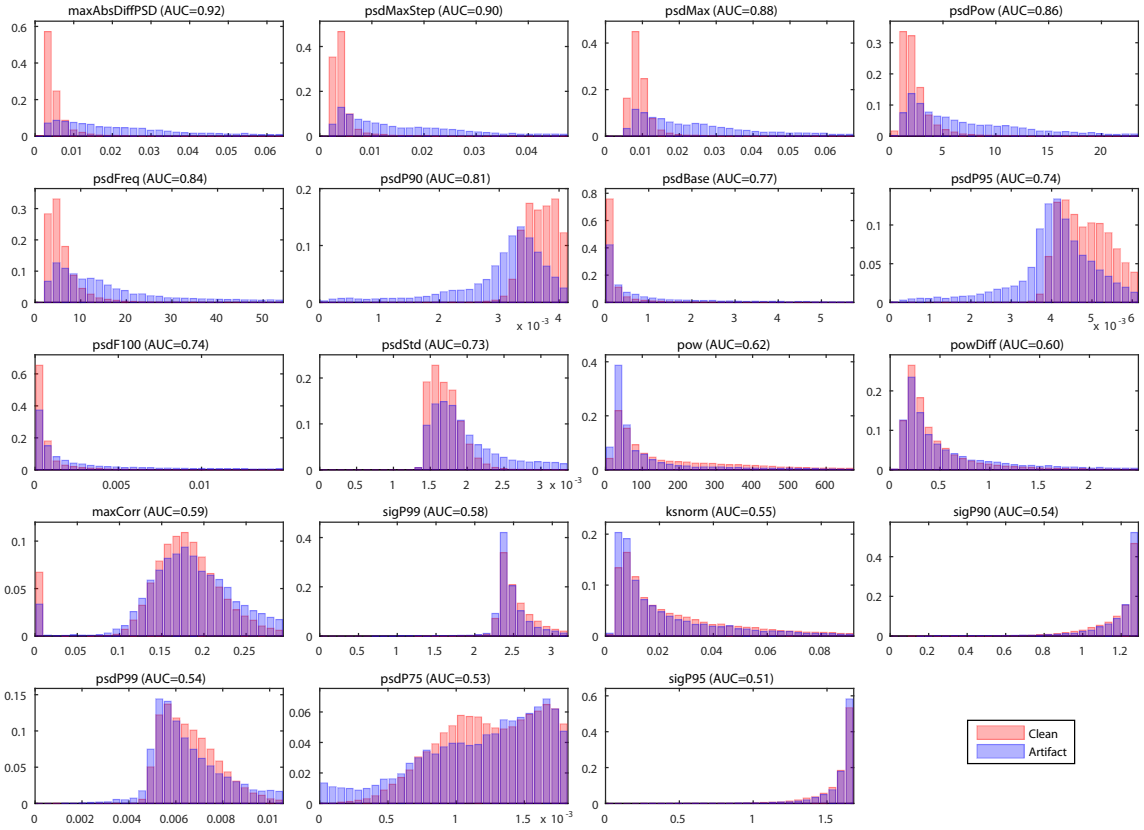

Figure 4.6: Histograms of all feature values on the Prague-CV database, sorted by area under receiver-operator characteristic (AUC) in descending order with artifacts in blue and clean signals in red. The similarity between histogram shapes (as well as between the feature definitions) suggests high inter-feature correlation, which was confirmed by the scatterplot matrix and computation of correlation coefficients. The classification methods used need to be chosen in order to handle high correlation within the feature set.

4.4.4 Classification results

The main classification performance evaluation was done according to the procedure described in Section 4.2.4, the process is shown graphically in Figure 4.3. After parameter selection, done on the Prague-CV cross-validation, performance of the final classifiers has been tested on the remaining four datasets. Results of the performance evaluation are presented in Table 4.6, the best performing classifier in each category (stationary segmentation, maxDiffPSD, tree and boosting) are shown in bold. The classification results show that the best performing classifiers: Bagging with 75 learners and the decision tree, achieved classification accuracy close to 90% on the cross-validation set and accuracy higher than 86% on three out of four of the unseen validation sets. Performance on the Olomouc set was markedly lower than on other validation sets for almost all classifier types — some comments on this matter can be found in the discussion section below.

The classification results were very good on three out of four unseen validation sets, while the performance on the Olomouc dataset dropped by up to 8% in comparison. As the ratio of different artifact types in this dataset is not different from e.g. data from Brno (see Figure 4.5 and Table 4.5), we may assume, that the reason for this performance drop is not in different artifact content of this dataset. Conversely, the inter-rater agreement on this dataset (Table 4.4) was lowest from all centers, which suggests more noise in the annotation and poorer classification performance is therefore to be expected. This may put into question the annotation methodology, however the good inter-rater concordance on the remaining two centers (Brno and Bratislava), unseen by the team during development of the rating methodology support its reliability. The character of artifact-affected signals differed slightly among centers, although we did not identify distinct problems in our artifact categories, that would lead to changes in the methodology.

The overall validation classification performance reached 87% in the case of the bagging classifier and almost as much for the decision tree classifier, achieving accuracy in the range between 87.0% and 89.4% on all centers except Olomouc, where the best performing decision Tree reached 80.9% only. On these centers, the performance of the two prominent classifiers was closely followed by the SVM classifier with linear discriminant classifier as a feature selector and linear kernel. The considerably simpler classifier *maxDiffPSD*, using threshold on a single spectral measure, also achieved a respectable overall accuracy between 80.2% and 86.2% on the three well-performing centers (82.7% overall), despite the markedly lower specificity and on the other hand higher sensitivity values on all sets. The chosen classification threshold, based on the Youden index on crossvalidation sample seems therefore slightly too low and therefore strict. This is further supported by the performance evaluation in Figure 4.7 a): a slightly higher accuracy would be achieved even on the crossvalidation set using higher detection threshold

at the cost of lower sensitivity. Based on these observations, the sigInspect automatic signal annotator was extended with the possibility to set arbitrary classification threshold for this classifier type.

As for the segmentation approaches *COV* and *SWT*, based on the original research by (Falkenberg et al. 2003; Aboy et al. 2006) and (Guarnizo et al. 2008), the results were by approximately 8-10% inferior to the best performing aforementioned methods. This may be attributed to the fact, that the methods only segment the signal at substantial change points and use no information about properties of clean signal nor of artifacts. In cases of stationary long-term artifacts, such as some cases of the *FREQ* type, a long and contaminated signal section may therefore be selected as the longest stationary component. This property is inherent to the unsupervised nature of both the original methods and their extended versions, presented in this chapter.

It may be noted, that the optimal parameters included the lower bound of available time-window lengths — 0.25 s — and also the lowest available aggregation threshold: each second was divided into four segments, presence of a single segment labeled by the classifier as artifact sufficed to label the whole second as artifact. The latter property was chosen in all cases also for longer windows (0.33 s and 0.5 s), which apparently provides the classifier with better ability to detect short-term events appearing within the one-second signal. The dependency of crossvalidation performance on detection threshold and window length for both methods can be found in Figure 4.7 b) and c). It can be noted that the performance was very close for all short windows — especially 0.25 s and 0.33 s — and using even shorter windows would most likely lead to very minor, if any, performance improvement.

The additional post-hoc crossvalidation procedure B), described in Section 4.2.4 consisted of repetitive classifier training on data from the original test set (Prague-TS, Bratislava, Brno, Olomouc) except one subject (leave-one-subject-out or LOSO) and validation on the remaining subject — a procedure which lead to 14 crossvalidation folds, corresponding to the total of 14 patients in all four original sets together. The aim was to identify possible benefit of training on multi-center dataset, instead of Prague-CV only. Mean values of classification results on the validation samples are presented in Table 4.7. Apparently, the classification accuracy increased slightly in almost all cases, especially on the Olomouc set, where the SVM, RobustBoost and other boosting classifiers experienced accuracy increase of up to 8.6 percentage points. It is notable that almost all specificity values increased in this classification scenario, usually at the cost of decreased sensitivity. This is to be expected due to higher diversity in artifact types in different centers. It can be seen that especially the RobustBoost classifier benefited most from this additional information and its classification accuracy was comparable to other boosting techniques. Please note, that in this scenario the parameters were optimized on the training data within each fold, which lead to varying classifier setting across folds. Thus, the results

may be a positively biased estimate of out-of-sample classification performance using the whole original test set as training data.

An additional point of view on the crossvalidation performance of different methods can be found in Figure 4.8 and suggests that the increase in mean performance especially of the COV segmentation technique was mainly due to lower number of missclassified validation set examples (note the bottom edge of the corresponding box, representing the 25th percentile). In contrast, the performance of the SWT method surprisingly slightly decreased. The overall classification performance on subjects from the Olomouc center was much better, compared to the main crossvalidation A scenario.

The final decision was to include four classifiers within the *sigInspect* toolbox: COV, maxDiffPSD, Tree and Bagging-based ensemble. The selection was mainly motivated by the good classification results, multiple classifier types were included to allow the users to choose the most appropriate method for their data processing scenario. Instructions for use and list of all available methods and parameters can be found in the *sigInspect* documentation.

4.4.5 Discussion

This section comments on the limitations of the presented study and provides additional discussion of the experimental resuls. One of the limitations stems for the composition of the training and validation set: all data were from DBS microelectrode trajectories, targetting the subthalamic nucleus. Despite this shortcoming, and the fact that all centers used the Medtronic Leadpoint targetting system, we believe that based on the presented results, the *sigInspect* tool may be of benefit to the research community and may be used for automatic or semi-automatic MER signal denoising.

Thanks to the extensive procedure for identification of appropriate artifact types for annotation, as well as for harmonization of the team of raters, the annotation reliability was satisfactory (achieving around 90% accuracy on the proofing sample, see Table 4.4) but still leaving a significant zone with unclear annotation. However, inspection of signal examples with low agreement showed mostly unclear cases where the artifact was either very weak and therefore questionable, or very short in time and easy to mistake for physiological spike. Both these cases are very hard to objectively distinguish under no ground-truth data, which is achievable only in laboratory conditions or computer simulations.

In our experience, the spectrogram was very helpful for revealing artifacts not easily visible in the time series plot (especially the *FREQ* type), as was the auditory inspection. Despite the use of these tools, our early experiments proved accurate identification of the exact artifact start and end time very challenging, mainly due to their gradual nature. Our annotation procedure thus used one-second segments, as did the presented classifiers. We believe, that a similar technique

could be used for shorter signal segments as well, although we find the one-second windows convenient for manual annotation. To date, we also do not possess a sufficient battery of MER signals with annotation on shorter segments, which could be used for classifier evaluation on shorter time windows.

The classification performance was evaluated using a strict crossvalidation scenario (A), in which all data from other centers were laid aside until the final testing stage at which all classifier parameters were already set and fixed. Therefore, the observed performance measures should represent an unbiased estimate of classification outcome on a novel, unseen dataset. Additionally, the post-hoc LOSO crossvalidation (B), which utilized data from the original testing sets for classifier training and parameter optimization, showed that a diverse training dataset may lead to a more robust classification at the cost of a decrease in sensitivity.

We may conclude, that the additional variability, introduced by data from other centers into the training dataset, lead to improved performance on the most poorly performing centers, while the overall performance increased slightly in all classifiers. This improvement was driven by large increases in specificity and counter-balanced by decreases in sensitivity. Generally speaking, training on a multi-center dataset lead to a more robust, although less sensitive classification performance, which we find desirable for a new *sigInspect* user with his own dataset. The classifiers included in the *sigInspect* toolbox were therefore trained using all datasets, presented in this study.

Despite the extremely high detection accuracies, presented by the authors of the segmentation approaches *COV* (Falkenberg et al. 2003; Aboy et al. 2006) and *SWT* (Guarnizo et al. 2008), which were almost 100% on simulated data, we show that the real-world performance of these methods in selecting clean signal segments may not be as outstanding and was superseded by all the other newly proposed classifiers by a relatively large margin in all settings.

All classifiers, presented in this chapter ignore the differences between artifact types. As not all distinct artifact types have negative impact in all data processing pipelines (e.g. weak *FREQ* artifacts do not necessarily affect results of studies on single-unit activity, even though they may impact detection of background noise level), such functionality might be useful. This decision was necessary due the low agreement on exact artifact type among the raters. To provide the user with better control of sensitivity to various classifier types, we have included the possibility to modify detection threshold of several methods, to conform to particular researchers experience and processing pipeline.

A close inspection of the samples misclassified by the proposed classifier revealed many borderline cases and erroneous annotation and we therefore believe that the classification performance, perceived by the end user of the sigInspect tool will be even better than our evaluation suggests. From our tests on additional long-term data, we believe that the system can fulfill the purpose of MER preprocessing and artifact identification tasks.

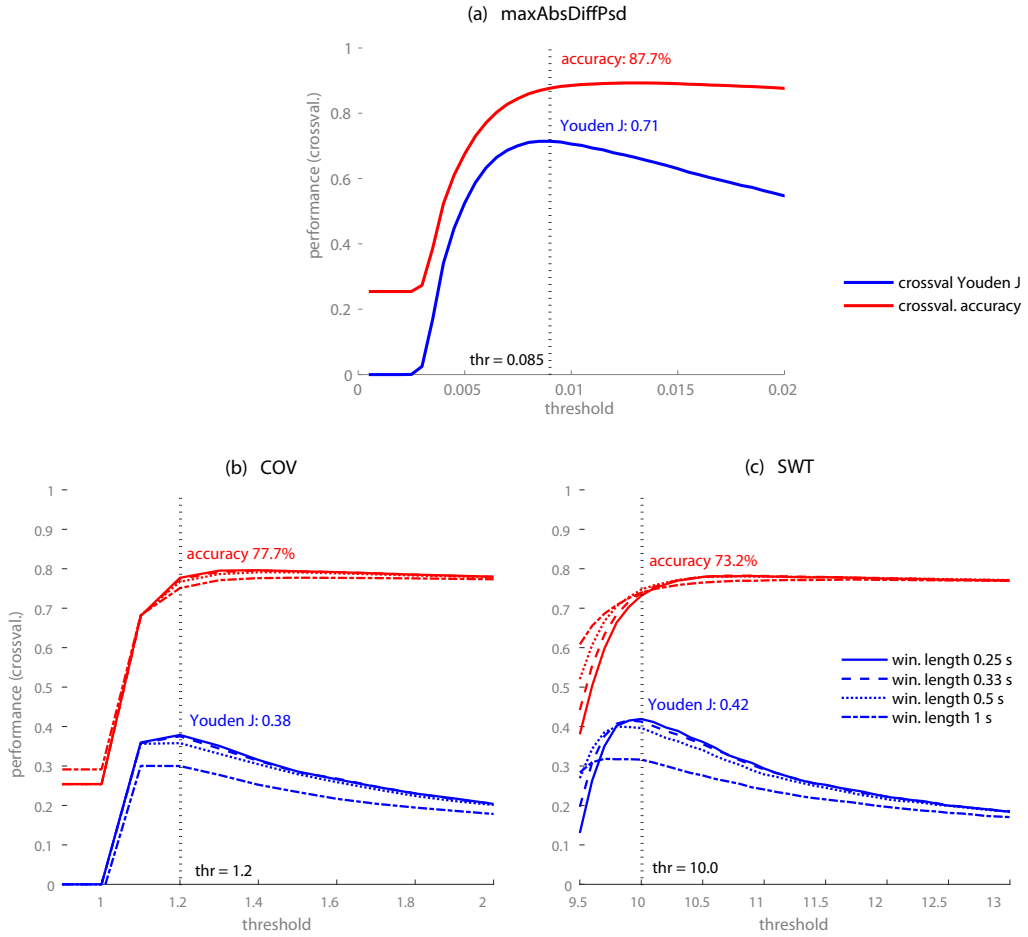

Figure 4.7: Evaluation of artifact detection using a) maximum difference from a clean sample spectrum *maxDiffPSD* b) extended *COV* method (Aboy et al. 2006) and c) extended *SWT* method (Guarnizo et al. 2008). Performance on the Prague-CV crossvalidation set is shown versus detection threshold for each method. For all methods the threshold which optimizes the J-statistic is shown as a vertical dotted line and achieves sub-optimal accuracy. A slightly higher accuracy could be achieved at the cost of decreased sensitivity.

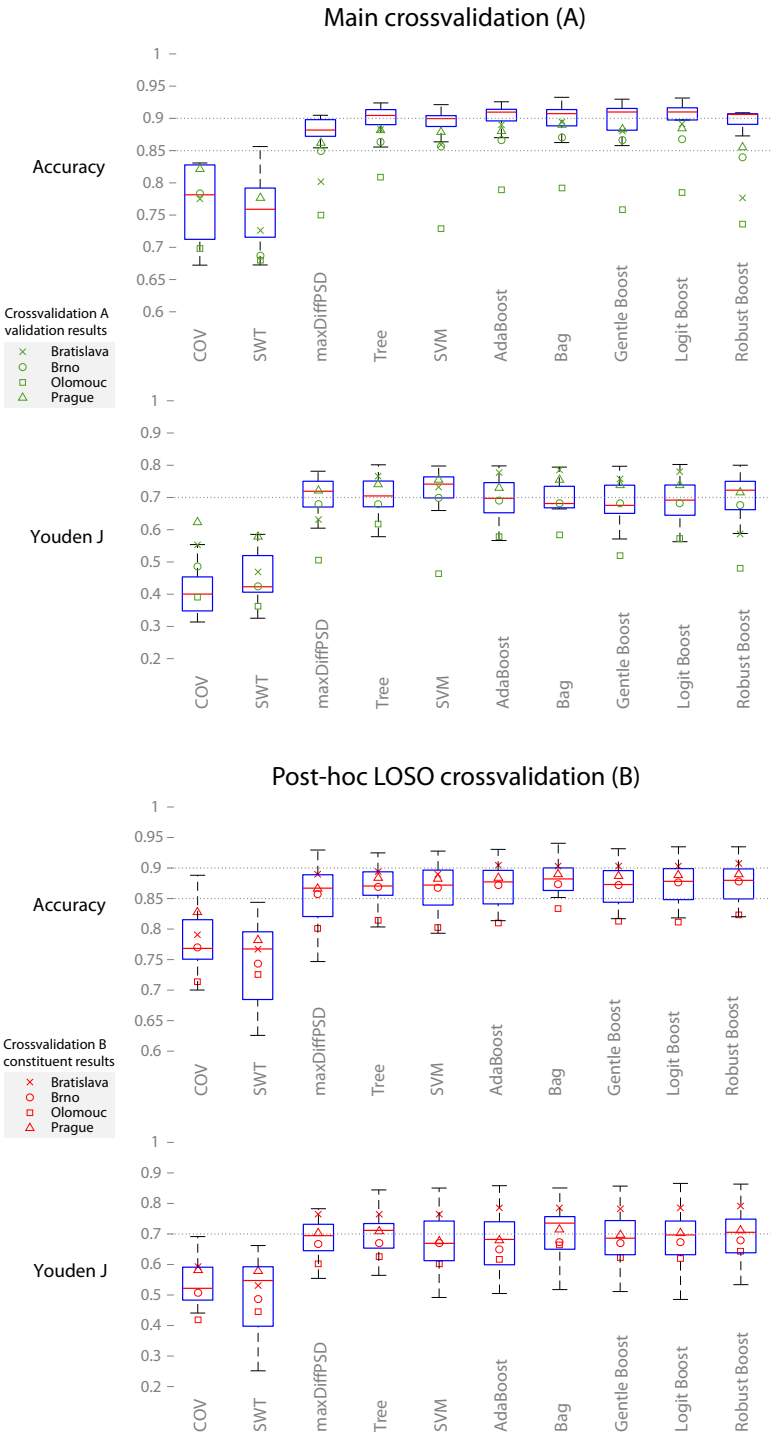

Figure 4.8: Boxplot of classification results in both classification scenarios as defined in Section 4.2.4: A) the main crossvalidation with parameter optimization on the Prague-CV set. The boxes represent crossvalidation performance across the ten folds, the green markers denote test performance on the remaining four unseen datasets. B) Post-hoc crossvalidation scenario, the four validation centers are used for training and evaluation in the leave-one-subject-out (LOSO) scenario with data from fourteen patients. The red markers show average crossvalidation performance on patients from each center.

4.5 Chapter conclusions

The chapter presents several novel methods for classification of exogenous artifacts in the MER signals and presents the *sigInspect* tool for general use in the research community. In the evaluations on manually labeled data from four different DBS centers, including more than 20h of MER data from 69 patients, the best performing methods showed high accuracy between 87% and 89% in three out of four centers. Considering reliability of the manual annotation itself — despite the extensive training and team synchronization undertaken — the classification results may be seen as very good, closely reaching the annotation accuracy. Additional analysis of misclassified samples showed many borderline cases and annotation errors, suggesting that the ability of the proposed classifiers to detect actual exogenous events in the MER data may be even better than our performance estimates suggest.

The *sigInspect* tool presented herein, allows automatic or semi-automatic MER signal annotation and can be used for various labeling and inspection tasks also on other single and multi-channel signals. The tool includes many configuration options and an API to access arbitrary data sources and databases. To the best of our knowledge, the presented solution is currently the only tool for detection of exogenous MER artifacts, available in the neuroscience community.

Table 4.6: Results of main classifier crossvalidation A and evaluation on independent test sets

classifier type	params	Prague-CV crossval.			Prague-TS test			Bratislava test			Brno test			Olomouc test			**Overall test**		
		Acc	Sens	Spec	**Acc**	Sens	Spec	**Acc**	Sens	Spec	**Acc**	Sens	Acc	**Acc**	Sens	Spec	**Acc**	Sens	Spec
COV	win. .25s, perc. 25%, thr 1.20	**77,7**	51,1	86,8	**82,1**	76,2	86,0	**77,6**	79,3	76,2	**78,4**	62,6	86,0	**69,9**	54,3	84,9	**78,1**	69,7	83,3
SWT	win. .25s, perc. 25%, thr 10.0	**73,2**	66,4	75,5	**77,7**	84,3	73,5	**72,7**	81,6	65,3	**68,7**	78,1	64,2	**68,0**	70,2	65,9	**72,3**	79,7	67,4
maxDiffPSD	threshold 0.0085	**87,7**	81,7	89,7	**86,2**	85,4	86,6	**80,2**	95,4	67,8	**84,9**	81,4	86,6	**75,0**	89,5	61,0	**82,7**	87,3	78,2
Tree	parent: 1000, leaf 250, deviance	**89,5**	76,3	94,0	**88,2**	81,4	92,6	**88,3**	88,1	88,4	**86,4**	77,1	90,8	**80,9**	77,7	83,9	**86,6**	81,3	89,8
SVM	feat.sel.: linear, thr 0.001, linear	**88,8**	83,3	90,6	**87,8**	87,1	88,3	**86,0**	92,9	80,3	**85,6**	83,3	86,7	**72,9**	92,0	54,5	**84,6**	88,1	81,1
AdaBoost	100 learners, learn. Rate 0.7	**89,8**	75,7	94,6	**88,0**	79,3	93,7	**89,0**	87,7	90,1	**86,6**	78,7	90,3	**78,9**	84,6	73,4	**86,6**	82,0	88,9
Bagging	75 learners	**90,0**	75,9	94,7	**89,0**	82,1	93,5	**89,4**	88,3	90,3	**87,0**	75,9	92,3	**79,1**	83,0	75,5	**87,1**	81,9	89,8
GentleBoost	250 learners, lrn. rate 0.1	**89,8**	75,8	94,6	**88,3**	80,7	93,2	**88,0**	87,6	88,2	**86,7**	76,7	91,4	**75,8**	85,4	66,6	**86,0**	81,9	87,7
LogitBoost	150 learners, lrn. rate 1	**90,0**	75,6	94,9	**88,4**	80,1	93,8	**89,1**	88,5	89,6	**86,8**	76,5	91,7	**78,5**	86,1	71,2	**86,7**	81,9	89,0
RobustBoost	10 learners, e. goal 0.2, e. marg. .1	**86,7**	83,2	87,9	**85,5**	87,0	84,6	**77,7**	96,3	62,3	**83,9**	83,6	84,1	**73,7**	91,4	56,6	**81,4**	89,0	74,8

Table 4.7: Results of post-hoc crossvalidation B (LOSO on validation centers).

classifier type	Prague-TS val.			Bratislava val.			Brno val.			Olomouc val.			**Overall val.**		
	Acc	Sens	Spec	**Acc**	Sens	Spec	**Acc**	Sens	Spec	**Acc**	Sens	Spec	**Acc**	Sens	Spec
COV	**82,7**	69,5	88,7	**79,0**	76,1	83,2	**77,0**	71,3	79,5	**71,3**	51,6	90,4	**78,4**	69,3	84,6
SWT	**78,2**	77,5	80,5	**76,7**	74,5	78,7	**74,3**	73,3	75,4	**72,5**	57,9	86,7	**75,8**	72,7	79,3
maxDiffPSD	**86,7**	81,6	88,6	**89,0**	84,4	92,1	**85,7**	77,2	89,6	**80,1**	74,9	85,3	**86,1**	80,0	89,4
Tree	**88,3**	79,0	91,9	**89,4**	85,2	91,4	**86,9**	75,2	91,9	**81,4**	81,7	81,0	**87,2**	79,8	90,3
SVM	**88,2**	75,0	92,7	**88,9**	86,5	90,1	**86,8**	77,5	89,4	**80,3**	80,1	80,1	**86,9**	79,5	89,3
AdaBoost	**88,4**	74,1	93,9	**90,4**	85,3	93,2	**87,2**	72,0	92,9	**81,0**	80,9	80,8	**87,5**	77,3	91,6
Bagging	**89,0**	78,8	92,7	**90,3**	87,4	91,3	**87,3**	75,7	91,5	**83,3**	84,1	82,3	**88,0**	80,8	90,5
GentleBoost	**88,6**	75,7	93,9	**90,3**	86,0	92,4	**87,2**	75,2	91,6	**81,3**	81,9	80,3	**87,6**	79,1	90,9
LogitBoost	**88,8**	76,7	93,5	**90,3**	85,8	92,8	**87,7**	74,5	92,7	**81,2**	81,2	80,8	**87,8**	79,0	91,3
RobustBoost	**88,9**	77,4	93,6	**90,8**	86,2	93,0	**87,8**	75,7	92,3	**82,3**	82,9	81,3	**88,1**	79,9	91,4

Chapter 5

Probabilistic model of DBS microelectrode trajectories

Chapter summary

This chapter presents results of our effort to create a probabilistic model of DBS microelectrode passes, that would be in further perspective suitable for fitting a three-dimensional anatomical STN model to multi-electrode recordings. Contrary to existing solutions using Bayes classifiers or Hidden Markov Models, the suggested model uses smooth state-transitions represented by sigmoid functions, which ensures flexible model structure in combination with general optimizers for parameter estimation and model fitting. The presented model can easily be extended with additional parameters and constraints, as can be found in Chapter 6.

In an evaluation on 260 trajectories from 61 patients, the model showed classification accuracy 90.0%, which was comparable to existing solutions. The evaluation proved the model successful in target identification and it can be concluded that its use for more complex tasks in the area of DBS planning and modeling is feasible.

The text of this chapter is an extended version of thepaper:

- **Bakstein, E.**, Sieger, T., Novak, D. and Jech, R. "Probabilistic Model of Neuronal Background Activity in Deep Brain Stimulation Trajectories". In: *Lecture Notes in Computer Science, Proceedings of the 7th international conference ITBAM 2016*, Springer Verlag.

5.1 Introduction

To obtain an accurate location information about surgical target in DBS surgery, a method called microrecording is employed in a vast majority of DBS centers (Abosch et al. 2013) (see Chapter 2 for introduction to the topic). In microrecording, a set of microelectrodes (tip diameter around 5 μm) is shifted through the brain and microelectrode EEG (also μEEG or MER) is recorded. The recorded signals are evaluated concurrently by a trained neurologist, who then identifies optimal position for the stimulation contacts. The evaluation is typically based on visual and auditory inspection of the signals, the main markers being neuronal firing pattern and especially amplitude of the neuronal background, which are higher in areas with higher neuron density — such as the STN. The accumulation of neurons in the STN is very high compared to the neighboring structures, which projects into the recorded signals as a sudden increase in the neuronal background activity as the electrode approaches the STN boundary, as well as appearance of rapidly spiking neurons once the electrode entered the nucleus. The former can be estimated by the root mean square (RMS) of the original signal (Moran et al. 2006; Zaidel et al. 2009), some authors also suggested signal with removed spikes or RMS of a band-pass filtered signal (Novak et al. 2007; Novak et al. 2011).

For a long time, efforts have been made to use machine learning models in place of the manual evaluation. This paper presents a probabilistic model of neuronal background activity along a microrecording trajectory, characterized by a normalized root-mean-square measure (NRMS). The suggested model is a logical extension of already existing models, which are summarized in the next section.

5.1.1 Existing models

Early models used the neuronal background level, estimated using the normalized root-mean-square of the signal as an input to Bayesian classifier (Moran et al. 2006) or discrete hidden Markov model (HMM) (Zaidel et al. 2009). These models included also the expected distance to target as an input, which utilizes the fact that the pre-surgical planning places the target (i.e. "depth 0") to a specific part of the STN. These models also used manual quantization or thresholding of the input parameters in order to achieve reasonably-sized discrete parametric space, that can be estimated from commonly-sized training datasets.

Extension to semi-markov models, including state duration (i.e. the length of nuclei pass) with continuous probability density function has been done by Taghva et al (Taghva 2011), but has been evaluated only on simulated data. Other researchers investigated features such as high-frequency component of the neuronal background (Novak et al. 2007) or multiple features including power spectral density, firing rate and noise level coupled with a rule-based classifier composed of cascaded thresholds (Cagnan et al. 2011). Support vector machine classifier on

multiple signal features (RMS, nonlinear energy, curve length, zero crossings, standard deviation and number of peaks) has been also implemented by Guillen et al (Guillen et al. 2011) with almost 100% accuracy.[1]

The authors of (Shamir et al. 2012) investigated the impact of recording length and density of recording depths on performance of an HMM and concluded that precision of a previously published HMM model (Zaidel et al. 2009) was approximately half of the between-position distance.

5.1.2 Proposed model

In this chapter, we present a model based on the neuronal background level, which can be used as a basis for fitting anatomical 3D model directly to the recorded μEEG activity along parallel trajectories. The presented variant is a one-dimensional proof of concept, intended to verify the idea and compare its properties to existing well-performing models.

Similarly to the hidden semi-markov models used in (Taghva 2011), our model uses parametric representation of input feature space – the NRMS values computed according to (Moran et al. 2006) but without quantization. Contrary to HMM, our model uses smooth state to state transitions, motivated by properties of electrical field of the STN, observed on the training data.

A derived model, based on the proposed approach, can be used to introduce other requirements such as the expected length of STN pass for given trajectory, based on a-priori information from surgical planning. Owing to the smooth state transitions, the model has also a smooth likelihood function (and gradient) and can be fitted using general purpose optimization algorithms. Thanks to this property, the structure of the model is very flexible and can be easily modified and extended. Moreover, the model theoretically allows classification with accuracy beyond the resolution of the measured data. However, this may not be the case practically due to noise in the μEEG signal and other measurement inaccuracies.

5.2 Methods

The probabilistic model, presented in this paper, is based on the assumption of different distribution of neuronal background level before, within and beyond the STN. Each of these distributions is represented parametrically and transitions between the consecutive distributions are modeled by the logistic sigmoid function (see section 5.2.2 below). In this section, we give

[1]The dataset in (Guillen et al. 2011) consisted of 52 signals from four patients only and it is not clear whether the validation sample was completely independent in terms of similarity of neighbor segments — see e.g. (Hammerla et al. 2015) for description of a similar problem.

overview of the proposed model, as well as of the data collection and pre-processing.

5.2.1 Data collection, annotation and pre-processing

The experimental dataset was collected during the standard surgical procedure of DBS implantation using a set of one to five tungsten microelectrodes, spaced 2 mm apart in a cross; the so-called Ben-gun configuration (Gross et al. 2006). The microelectrode signals were recorded at each 5 mm along the trajectory using the Leadpoint recording system (Medtronic, MN), sampled at 24 kHz, band-pass filtered in the range 500–5000 Hz and stored for offline processing. Annotation of nucleus at each position was done manually by an expert neurologist, based on visual and auditory inspection of the recorded signal.

To reduce the effect of motion-induced artifacts, we divided each signal into $1/3$ s windows and selected the longest stationary component using the method presented in (Bakstein et al. 2015), which is an extension of method previously presented in (Aboy et al. 2006). Parameters of the method (detection threshold and window length) were selected in order to achieve best accuracy on a training database. This method was chosen in order to obtain at least some segment of each signal, even though it may contain electromagnetic and other interference, which would be marked as signal artifact by the stricter spectral method, presented in (Bakstein et al. 2015) or other methods from Chapter 4.

5.2.2 Electric field of the STN

In orded to obtain estimate of the neuronal background activity level, we calculated the root-mean-square (RMS) of the stationary portion of each signal. In accordance with (Moran et al. 2006), we computed the normalized RMS of the signal (NRMS) by dividing feature values of the whole trajectory by mean RMS values of the first 5 positions (which are assumed non-STN in a majority of recordings). Our observations are well in line with observations of P. Novak (Novak et al. 2007; Novak et al. 2011) - see the Figure 5.1 for example electrode passes from the training data. Additionally, we normalized the 90th percentile of each NRMS trajectory to 3 in order to limit NRMS variability in the STN.

Observations of NRMS values before, within and after the STN confirmed different distribution in each part. After comparing shapes of the probability density functions in each region, as well as values of likelihood of normal and log-normal distribution, we chose to model the NRMS values in each part by the best-fitting log-normal distribution. Further observations proved the NRMS-distribution to be relatively stable in the regions before, after and within the STN, as can be seen from the Figure 5.2. The Figure 5.3 shows comparison of probability density functions of the original data, as well as of the resulting model parameters.

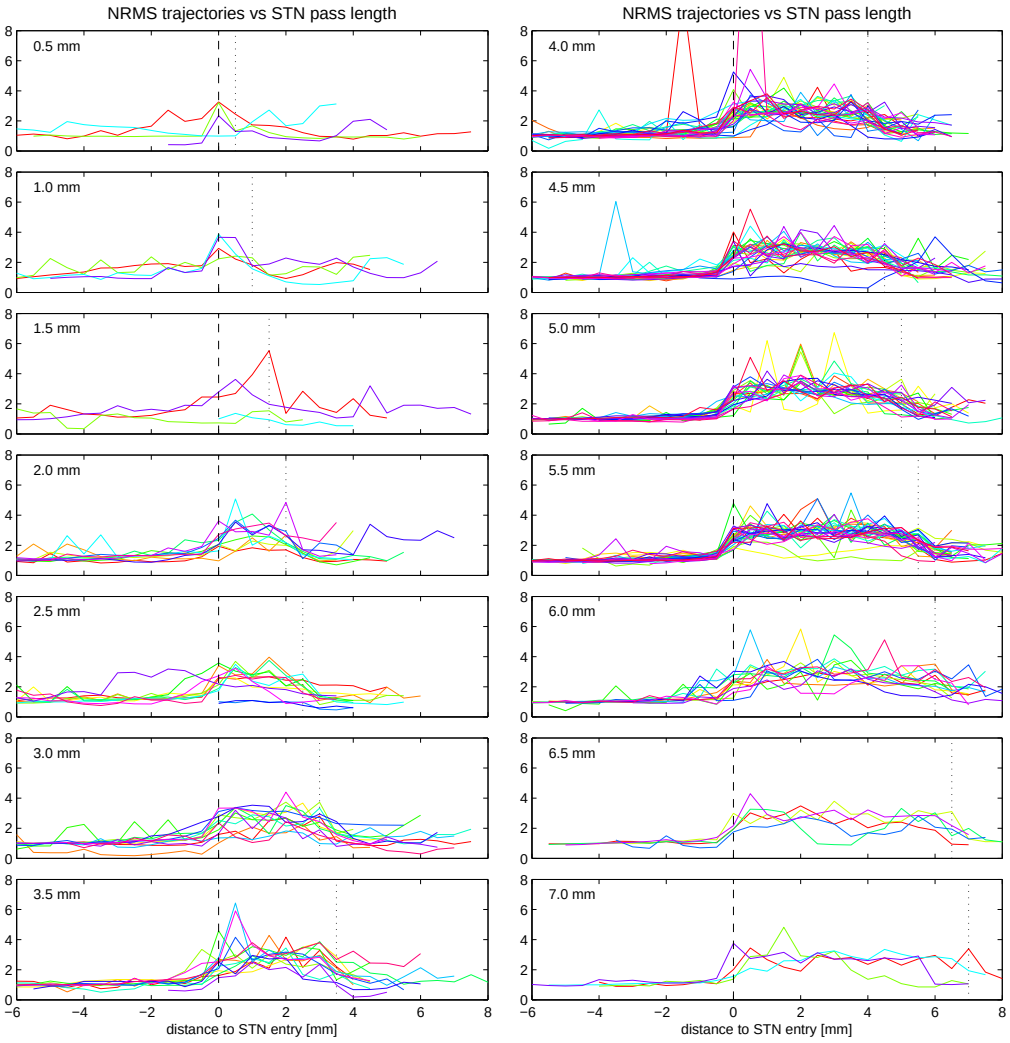

Figure 5.1: NRMS values along STN passes, grouped by STN pass length. STN entry and exit points are denoted by the dashed and dotted lines, respectively.

Further explorative analysis was aimed at the shape of NRMS transition. Figure 5.4 presents NRMS training data, aligned around STN entry and exit, mean value for each distance to the transition and the sigmoid logistic function we chose to model the transition as a result.

5.2.3 Parametric model of STN background activity

Model structure

The proposed model of background activity along the DBS trajectory consists of probability density of the NRMS measure in the three different regions. These can be seen as continuous emission probabilities in three hidden states of an HMM. Contrary to an HMM, the proposed model uses no discrete state transitions that could be represented by a transition matrix, but

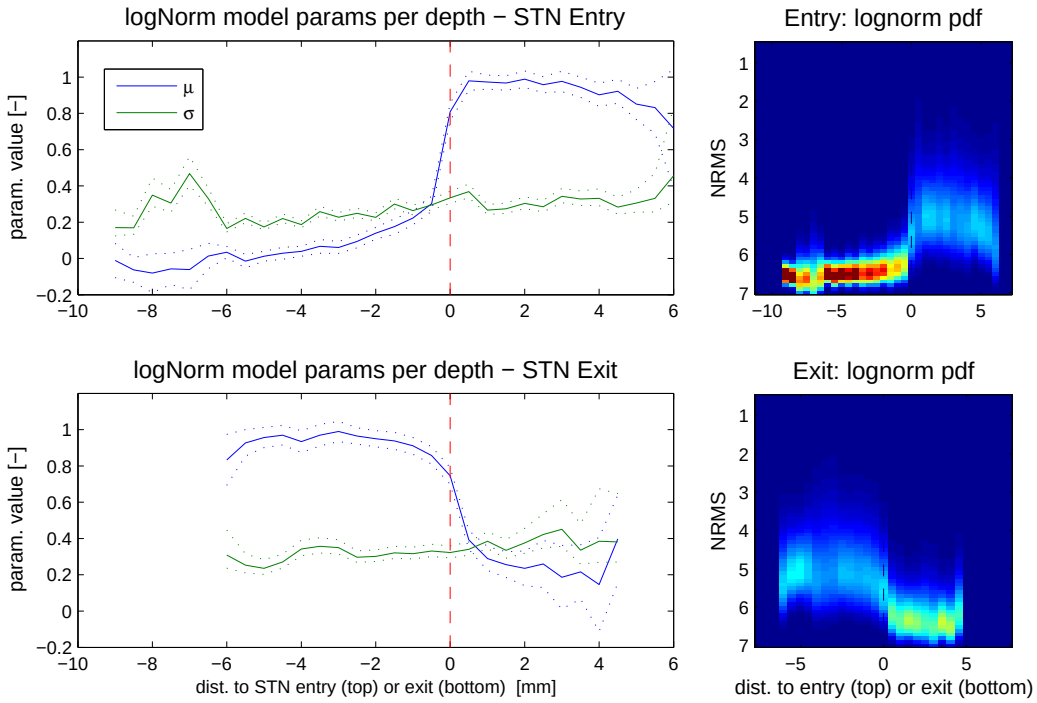

Figure 5.2: Resulting parameters of the log-normal distribution when fitted to all available NRMS data at a specific distance to STN entry (top) or exit (bottom) with 95% confidence intervals. It can be noted that the values within each region are relatively stable, with sharp rise around STN boundaries. Notice the slight slope in μ parameter before the STN entry; it can be seen that this phenomenon has relatively mild effect on the resulting probability densities on the right. The right panel shows probability density function of the fitted logNormal distribution as a heatmap — see Figure 5.3 for details.

uses smooth state transitions, represented by sigmoid (or logistic) functions. Due to that, standard evaluation methods used for HMM, such as the Viterbi algorithm, can not be used and are replaced by general constrained optimization.

The general idea of the proposed model is based on the following reasoning: one of the most obvious features, distinguishing DBS target structure in the μEEG — in particular the STN — is signal power, represented here by signal NRMS. Based on our observations on training trajectories (see Section 5.2.2), as well as previous works (e.g. (Novak et al. 2007; Novak et al. 2011)), we assume different probability distribution of NRMS values in the areas before, within and beyond the STN and use the log-normal distribution as a model for the NRMS values in each area. Parameters of the log-normal model are estimated from labeled training data during the training phase.

In common settings, the μEEG signals are recorded at discrete depth steps (in our case every

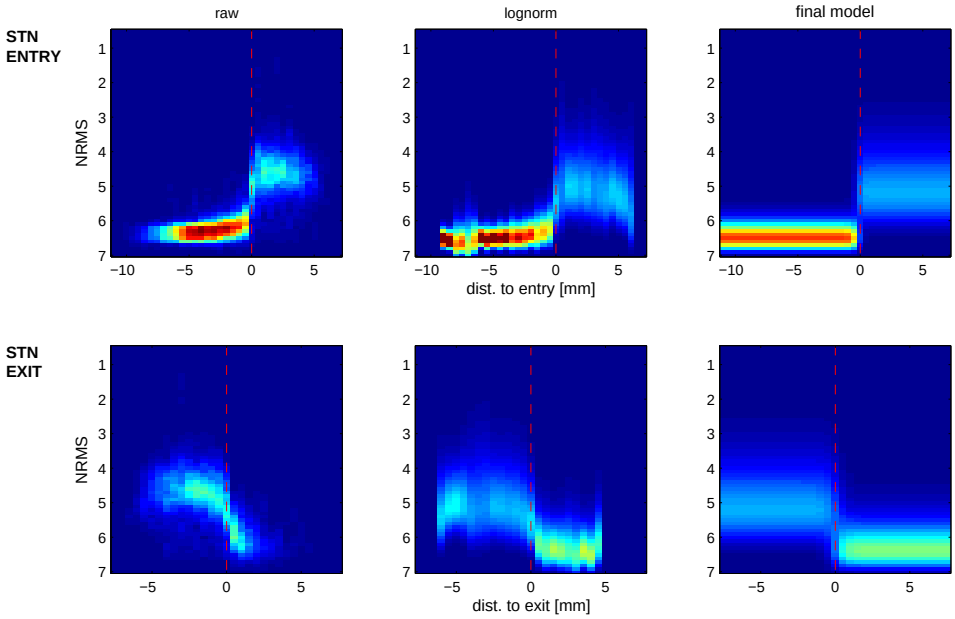

Figure 5.3: Probability density functions per given distance to STN entry (top) and exit (bottom) shown in the form of a heatmap. The left panel shows normalized 2D histogram of the original NRMS data. The middle column shows situation where log-normal model was fitted for each distance to entry/exit point in .5mm increments and the right panel shows probability density function of the resulting model: each region fitted with a single log-normal model with sigmoid transition functions.

0.5 mm). The task is therefore to classify signals, recorded at each position, to a correct class (i.e. identify the STN). We assume that the electrode can pass through the STN at most once and the trajectory can thus be divided into three consistent segments by two boundary points: STN entry and STN exit. In the evaluation phase we find optimal STN entry and exit points by maximizing the joint likelihood of the observed NRMS values along given trajectory with respect to the previously identified probability distributions. Simply put, the values before the assumed STN entry should be close to the expected value of the distribution before the STN, the values within the assumed STN should be close to the expected value of the distribution within STN and accordingly for the area beyond STN.

In order to increase theoretical precision of the model, as well as to improve its algebraic properties[2], we add smooth state transitions, modeled using logistic sigmoid functions. This approach also seems to be well in alignment with the observed statistical properties of NRMS values around STN boundary points — as can be seen in Fig. 5.4. The result of this addition

[2]Smooth state transitions using logistic sigmoid functions lead to smooth gradient and the resulting model is therefore easier to optimize.

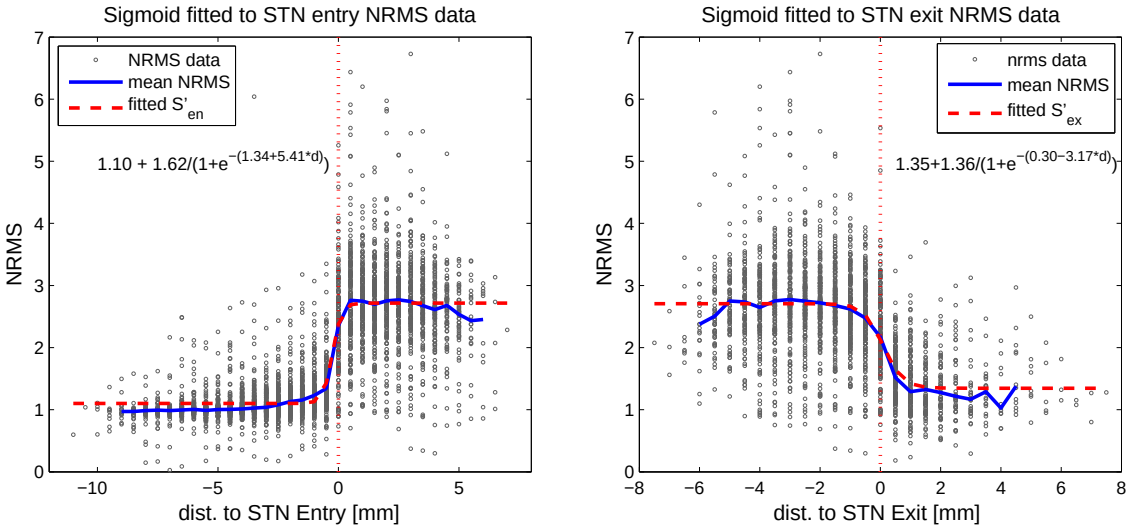

Figure 5.4: NRMS values around STN entry and exit points (depth 0 on the x axis) from a set of training trajectories. The blue line represents mean NRMS value for each distance, the red dashed line shows fitted sigmoid functions S'_{en} and S'_{ex}, used to model STN entry and exit transitions, with parameters corresponding to the inlaid formula.

is that rather than belonging to one particular state, each data point along the trajectory is assumed to be a partial member of all three states. Membership coefficients c_{pre}, c_{STN} and c_{post} of this combination are given by the sigmoid functions and depend on distance of given point from STN entry and exit. Illustration of the weighting can be found in Fig. 5.5.

In this paper, we present two variants of the model: i) the basic *flex1*, based solely on the NRMS measure and ii) extended model *flex2*, which adds a-priori distribution of expected STN entry and exit depths. The following sections provide formal definition of the model, as well as the training and evaluation procedure.

Training phase

Supervised model training is performed on NRMS feature values $x_i \in \{x_1, x_2, ..., x_N\}$, extracted from MER data recorded at N recording positions at depths $d_i \in \{d_1, d_2, ..., d_N\}$. Manual expert annotation is provided for each recording position, labeling the signal as either *stn* or *other*. STN entry position i_{en} and exit depth i_{ex} is defined as index of the first and last occurence of *stn* label from the start of the trajectory. Trajectory is then divided into three parts; i) before the STN with indices $I_{pre} = \langle 1, i_{en} - 1 \rangle$, ii) within the STN $I_{stn} = \langle i_{en}, i_{ex} \rangle$ and iii) after the STN $I_{post} = \langle i_{ex} + 1, N \rangle$. Two groups of parameters are fitted during the training phase:

i) Parameters of the log-normal probability distribution of NRMS feature values before the STN ($\boldsymbol{\theta}_{pre} = \{\hat{\sigma}_{pre}, \hat{\mu}_{pre}\}$), within the STN ($\boldsymbol{\theta}_{stn}$) and after the STN ($\boldsymbol{\theta}_{post}$), where $\hat{\mu}$ and

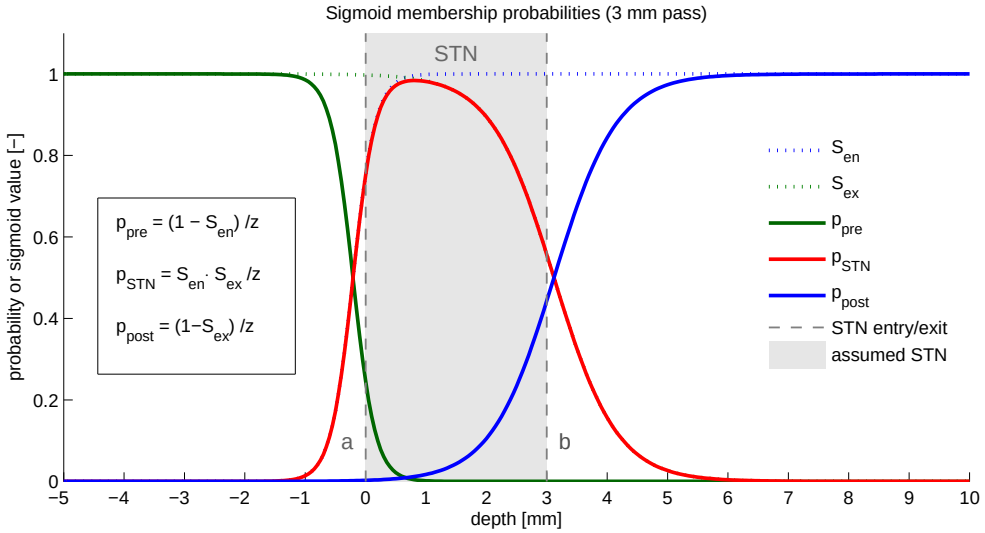

Figure 5.5: Illustration of sigmoid transition functions S_{en} and S_{ex} and their application to the joint likelihood function from Eq. 5.10: each observed data point is assumed to be a partial member of all three hidden states. Probability density functions corresponding to each state are weighted using the membership probabilities $p_{pre}(i) = p(d_i \in pre|a, b, \mathbf{\Theta})$, $p_{STN}(i) = p(d_i \in STN|a, b, \mathbf{\Theta})$ and $p_{post}(i) = p(d_i \in post|a, b, \mathbf{\Theta})$ which are dependent on distance from the hypothetical STN entry and exit points a and b. The $z(i) = z_i$ is normalization coefficient - see Eq. 5.12 and Eq. 5.15 for details.

$\hat{\sigma}$ are maximum-likelihood estimates of location and scale parameters of the respective log-normal distribution, computed in standard way according to

$$\hat{\mu}_{pre} = \frac{\sum_{i \in I_{pre}} \ln(x_i)}{n_{pre}} \tag{5.1}$$

$$\hat{\sigma}_{pre} = \sqrt{\frac{\sum_{i \in I_{pre}} \left(\ln(x_i) - \hat{\mu}_{pre}\right)^2}{n_{pre}}} \tag{5.2}$$

where $n_{pre} = |I_{pre}|$, i.e. the number of positions with given label. Parameters for *stn* and *post* labels are computed accordingly on samples from the I_{stn} and I_{post} sets.

ii) Parameters defining the shape of the sigmoid transition functions at STN entry (β^0_{en} and β^1_{en}) and exit (β^0_{ex} and β^1_{ex}}). Here, the parameter β^0 represents shift and β^1 steepness of the respective logistic sigmoid function, defined as

$$S'_{en}(d_i) = \alpha^0_{en} + \alpha^1_{en} \cdot \left(1 + \exp -(\beta^0_{en} + \beta^1_{en}(d_i - d_{en}))\right)^{-1} \tag{5.3}$$

for STN entry and

$$S'_{ex}(d_i) = \alpha^0_{ex} + \alpha^1_{ex} \cdot \left(1 + \exp -(\beta^0_{ex} + \beta^1_{ex}(d_i - d_{ex}))\right)^{-1} \tag{5.4}$$

for STN exit, where d_{en} is STN entry depth and d_{ex} STN exit depth. The additional parameters α^0 (shift along the y axis) and α^1 (scaling factor) serve to provide sufficient degrees of freedom to achieve appropriate fit. However, these parameters are not part of the model and are not stored as both are replaced by the log-normal probability density functions modeling the NRMS values in the respective area. Note that contrary to shifted and scaled functions S'_{en} and S'_{ex} fitted during the training phase, standard logistic functions S_{en} and S_{ex} from Eq. 5.13 and 5.14 are used during evaluation.

Fitting can be done using general purpose optimization function minimizing mean square error on all training data at once, according to:

$$\underset{\alpha^0_{en},\alpha^1_{en},\beta^0_{en},\beta^1_{en}}{arg\,\min} \sum_{i\in I_{pre},I_{stn}} \left(S'_{en}(d_i,\alpha^0_{en},\alpha^1_{en},\beta^0_{en},\beta^1_{en}) - x_i\right)^2 \tag{5.5}$$

and similarly for S'_{ex}. Only data labeled as *pre* and *stn* are used to fit parameters of S'_{en} and data labeled as *stn* and *post* are used to fit S'_{ex}. Initial parameters are set to $\left[\alpha^0_{en},\alpha^1_{en},\beta^0_{en},\beta^1_{en}\right] = [1,1,0,1]$ and $\left[\alpha^0_{ex},\alpha^1_{ex},\beta^0_{ex},\beta^1_{ex}\right] = [1,1,0,-1]$

The trained model is then completely characterized by parameter vector $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_{pre},\boldsymbol{\theta}_{stn},\boldsymbol{\theta}_{post},\beta^0_{en},\beta^1_{en},\beta^0_{ex},\beta^1_{ex}\}$, encompassing both log-normal emission probabilities and steepness and shift parameters of the sigmoid transition functions. If more trajectories are available for training, both parameter groups are estimated using all training data at once, given that appropriate labels and STN entry and exit depths are applied for each trajectory separately.

Extended model

The presented model structure uses no prior information about expected STN entry and exit depths. It is possible to modify the model by adding empirical distribution of expected entry and exit depths, modeled using the normal distribution $p_a = N(\mu_a,\sigma_a)$ and $p_b = N(\mu_b,\sigma_b)$. The parameters can be estimated using the standard maximum likelihood estimates of mean and standard deviation. This will lead to addition of four parameters. We will denote the extended parameter vector $\boldsymbol{\Theta}'$, the extended model is then nicknamed *flex2* in the results section.

Model evaluation

To evaluate the model on a vector of $\boldsymbol{x} = \{x_1,...,x_N\}$, measured on a particular trajectory at corresponding depths $\boldsymbol{d} = \{d_1,...,d_N\}$, we hold the model parameters $\boldsymbol{\Theta}$ fixed and use the maximum likelihood parameter estimation procedure to obtain the optimal posterior STN entry and exit parameters a and b. The likelihood function is defined as

$$\{a^*,b^*\} = \underset{a,b}{arg\,\max}\ \mathcal{L}(a,b|\boldsymbol{x},\boldsymbol{d},\boldsymbol{\Theta}) = \underset{a,b}{arg\,\max}\ L(\{\boldsymbol{x},\boldsymbol{d}\}|a,b,\Theta) \tag{5.6}$$

where $\mathcal{L}$ is likelihood of given the observation vector $\boldsymbol{x}$ at depths $\boldsymbol{d}$ and model parameters $\boldsymbol{\Theta}$, which can be calculated from L: the joint probability of observation vector $\boldsymbol{x}$ given model parameters. By naïvely assuming that observations from $\boldsymbol{x}$ are IID[3], we get

$$L(\{\boldsymbol{x}, \boldsymbol{d}\}|a, b, \boldsymbol{\Theta}) = \prod_{i=1}^{N} L(\{x_i, d_i\}|a, b, \boldsymbol{\Theta}) \tag{5.7}$$

and in order to avoid numerical underflow issues, we may shift to maximizing the negative log-likelihood function, which yields the same results for parameter a and b values and is computed according to:

$$\{a^*, b^*\} = \underset{a,b}{arg\,min} \sum_{i=1}^{N} -\ln(L(\{x_i, d_i\}|a, b, \boldsymbol{\Theta})) \tag{5.8}$$

The joint likelihood for position i at fixed values of STN entry and exit depths a and b and all three possible states (*pre*, *STN* and *post*) is given by:

$$\begin{aligned} L(\{x_i, d_i\}|a, b, \boldsymbol{\Theta}) &= p(\{x_i, d_i\}|a, b, \boldsymbol{\Theta}) = \\ &= p(x_i, d_i \in pre|a, b, \boldsymbol{\Theta}) \\ &\quad + p(x_i, d_i \in STN|a, b, \boldsymbol{\Theta}) \\ &\quad + p(x_i, d_i \in post|a, b, \boldsymbol{\Theta}) \end{aligned} \tag{5.9}$$

By expanding the probabilities in Eq. 5.9 using the Bayes' theorem, we get

$$\begin{aligned} L(\{x_i, d_i\}|a, b, \boldsymbol{\Theta}) &= p(x_i|d_i \in pre, \boldsymbol{\Theta}) \cdot p(d_i \in pre|a, b, \boldsymbol{\Theta}) \\ &\quad + p(x_i|d_i \in STN, \boldsymbol{\Theta}) \cdot p(d_i \in STN|a, b, \boldsymbol{\Theta}) \\ &\quad + p(x_i|d_i \in post, \boldsymbol{\Theta}) \cdot p(d_i \in post|a, b, \boldsymbol{\Theta}) \end{aligned} \tag{5.10}$$

where the probability $p(x_i|d_i \in pre, \boldsymbol{\Theta})$ represents the emmission probability in state *pre* and is computed using the standard probability density function of the log-normal distribution in the area before STN:

$$p(x_i|d_i \in pre, \boldsymbol{\Theta}) = \frac{1}{x_i \hat{\sigma}_{pre} \sqrt{2\pi}} \quad \exp - \frac{(\ln(x_i) - \hat{\mu}_{pre})^2}{2\hat{\sigma}_{pre}^2}, \tag{5.11}$$

using parameters of the log-normal distribution $\hat{\mu}_{pre}$ and $\hat{\sigma}_{pre}$, obtained in the training phase according to Eq. 5.1 and Eq. 5.2 respectively. The probabilities $p(x_i|STN, \boldsymbol{\Theta})$ and $p(x_i|post, \boldsymbol{\Theta})$ for NRMS distribution inside and beyond the STN are computed accordingly.

The class membership probabilities $p(pre|a, b, \boldsymbol{\Theta})$ from Eq. 5.10 (similarly for states *STN* and *post*) depend on the distance between depth d_i and currently assumed STN borders a and b and are computed from the sigmoid transition functions as follows:

$$\begin{aligned} p(d_i \in pre|a, b, \boldsymbol{\Theta}) &= (1 - S_{en}(d_i, a|\boldsymbol{\Theta}))/z_i \\ p(d_i \in STN|a, b, \boldsymbol{\Theta}) &= S_{en}(d_i, a|\boldsymbol{\Theta}) \cdot S_{ex}(d_i, b|\boldsymbol{\Theta})/z_i \\ p(d_i \in post|a, b, \boldsymbol{\Theta}) &= (1 - S_{ex}(d_i, b|\boldsymbol{\Theta}))/z_i \end{aligned} \tag{5.12}$$

[3]independent, identically distributed

using the sigmoid transition functions S_{en} and S_{ex}:

$$S_{en}(d_i) = \left(1 + \exp -(\beta^0_{en} + \beta^1_{en}(a - d_i))\right)^{-1} \tag{5.13}$$

for STN entry and equivalently

$$S_{ex}(d_i) = \left(1 + \exp -(\beta^0_{ex} + \beta^1_{ex}(b - d_i))\right)^{-1} \tag{5.14}$$

for STN exit. The z_i in Eq. 5.12 is a normalization coefficient ensuring that the class membership probabilities add to one under all circumstances[4]:

$$z_i = (1 - S_{en}(d_i, a|\mathbf{\Theta})) + S_{en}(d_i, a|\mathbf{\Theta}) \cdot S_{ex}(d_i, b|\mathbf{\Theta}) + (1 - S_{ex}(d_i, b|\mathbf{\Theta})). \tag{5.15}$$

In case of the **extended model** *flex2*, the minimization will take the following form:

$$\{a, b\} = \underset{a,b}{arg\,min} \left[\sum_{i=1}^{N} \left(-\ln(L(x_i, d_i|a, b, \mathbf{\Theta})) - \lambda ln(p_a(a|\mathbf{\Theta'}) \cdot p_b(b|\mathbf{\Theta'}))\right)\right] \tag{5.16}$$

where the summation of $L(x_i, d_i|a, b, \mathbf{\Theta})$ is the same as in Eq. (5.8) and the new $p_a(a|\mathbf{\Theta'})$ and $p_b(b|\mathbf{\Theta'})$ are probabilities of STN entry at depth a and exit at depth b, computed from the normal probability density function

$$p_a(a|\mathbf{\Theta'}) = \frac{1}{\sigma_a\sqrt{2\pi}} \quad \exp -\frac{(a - \mu_a)^2}{2\sigma_a^2} \tag{5.17}$$

and represent the prior probability of STN entry at depth a and exit at depth b. The parameter λ can be used to assign more/less importance to the a-priori depth distribution, compared to the observation-based likelihood element. In case of the presented results, we set the value of $\lambda = 1.75$ which optimized train-set accuracy. Dependency of the observed train and validation set accuracy can be found in Fig. 5.7

As this process can be vectorized and the parametric space is only two-dimensional and bounded, standard optimization algorithms with empirical gradient can be used to search for optimal parameters. In our case, we used constrained optimization with conditions requiring that $a \leq b$ (the entry depth a is lower or equal to exit depth b), $a \geq d_1$ and $b \leq d_N$ (entry and exit depths must be in the range of the data).

The parametric space may contain local optima (depending on the shape of NRMS values along given trajectory) and it is therefore very useful to provide reasonable initialization of a and b. In our implementation, the initialization was set as the mean entry and exit depths from the training data: μ_a and μ_b[5]. Note that both a and b are real numbers and are not restricted to the set of actually measured depths.

[4]Value of this normalization coefficient will however be close to one in most situations and reaches around 1.2 in the extreme case when $a = b$ using sigmoid parameters from Fig. 5.4.

[5]In the case with no entry/exit depth distribution, the initial parameters were set as the middle of the trajectory for a and the 3/4 of the trajectory for b

5.2.4 Crossvalidation

To evaluate the proposed model on real data and compare its classification ability against existing models, we evaluated the model in a 20-fold crossvalidation: in each fold, 5% of available trajectories were left out for validation, while the remaining data were used for estimation of model parameters. This lead to 20 sets of error measures for each classifier which were than averaged to obtain final estimates. Larger number of crossvalidation folds was chosen in order to obtain better estimate of error variability on different validation datasets.

The models compared were i) Bayes classifier from (Moran et al. 2006) based on discrete joint probability distribution of NRMS and depth and an ii) HMM model, based on the same discrete probability distribution (used as emission probabilities), with transition probabilities estimated from the training data in a standard way and two variants of the proposed model: iii)*flex1*, based solely on NRMS and iv) *flex2* with distribution of entry and exit depths.

5.3 Experimental results

5.3.1 Data summary

In total, we collected 6576 signals from 260 electrode passes in 117 DBS trajectories in 61 patients. Length of recorded signals was 10s. After discarding non-stationary signal segments, the mean length of raw signal segment that entered the NRMS calculation was $8.76s$ (median $9.67s$). In each crossvalidation fold, 13 electrode passes were used for validation, while the remaining 247 were used for training.

5.3.2 Classification results and discussion

Mean values of classification sensitivity, specificity and accuracy are presented in Table 5.1, while distribution of these error measures on the 20 validation sets can be found in Fig. 5.6. Even though the results of all methods were very similar (as can be seen especially in Fig. 5.6), the highest mean test accuracy was achieved by the *hmm* model – 90.2%, closely followed by the *flex2* model with 90.0%. Both models were also best in terms of specificity, while the best validation set sensitivity was achieved by the *hmm* and *bayes* classifiers.

Comparing two variants of the proposed method, the *flex2* model with entry depth distribution achieved better results than the NRMS-only variant *flex1*. The latter model tended to converge to local optima on trajectories with high noise level or non-standard NRMS shape.

Table 5.1: Classification results (error measures from 20-fold crossvalidation) comparing results of Bayes classifier (Moran et al. 2006) (*bayes*), Hidden Markov model (*hmm*), suggested model based solely on the NRMS (*flex1*) and extended model with distribution of STN entry and exit depth (*flex2*). See also Fig.5.6.

	TRAIN			TEST		
	accuracy	sensitivity	specificity	accuracy	sensitivity	specificity
bayes	90.4	84.1	94.1	89.0	82.5	92.8
hmm	91.3	83.8	95.7	90.2	83.1	94.3
flex1	88.5	80.9	92.9	88.0	80.6	92.2
flex2	90.1	83.2	94.1	90.0	83.1	94.1

5.3.3 Fitting of individual trajectories

Apart from the overall results, we also evaluated results on individual trajectories. The *bayes* model, which from definition put no constraints on the resulting label vector, was capable of classifying non-consecutive trajectories (interrupted STN labels) — this may have lead to the rather high sensitivity on the training data. As for the proposed models, the *flex1* NRMS-only variant tended to fit zero-length STN near the end of the trajectory in cases of non-standard STN passes where the NRMS did not exhibit the standard low–high–low profile or contained strong local peaks. The addition of entry and exit depth distribution in the *flex2* model variant reduced this problem and lead to improved classification accuracy.

An example of a successful STN classification on a typical trajectory using the *flex1* model can be seen in Fig. 5.8, while the corresponding negative log-likelihood function from Eq. 5.10 can be seen in Fig. 5.9. Note that the log-likelihood function is defined only for $a \leq b$. In the case of the *flex2* model, the values of the likelihood function around the a-priori expected entry and exit depth are further reduced by the additional component in Eq. 5.16, which increases the performance especially in cases with high noise in NRMS values.

5.4 Discussion

The presented model achieved comparable accuracy to existing approaches, represented by bayesian classifiers (Moran et al. 2006) and HMM (Zaidel et al. 2009). The results of HMM and hidden semi-markov models, presented by Taghva et al (Taghva 2011) were much superior, but were evaluated on simulated data only. In summary, the presented extended model (*flex2*) achieved mean classification accuracy 90.0%, sensitivity 83.1% and specificity 94.1% on the test set. As seen from the heavy overlap of different method's results, clearly visible in Fig. 5.6,

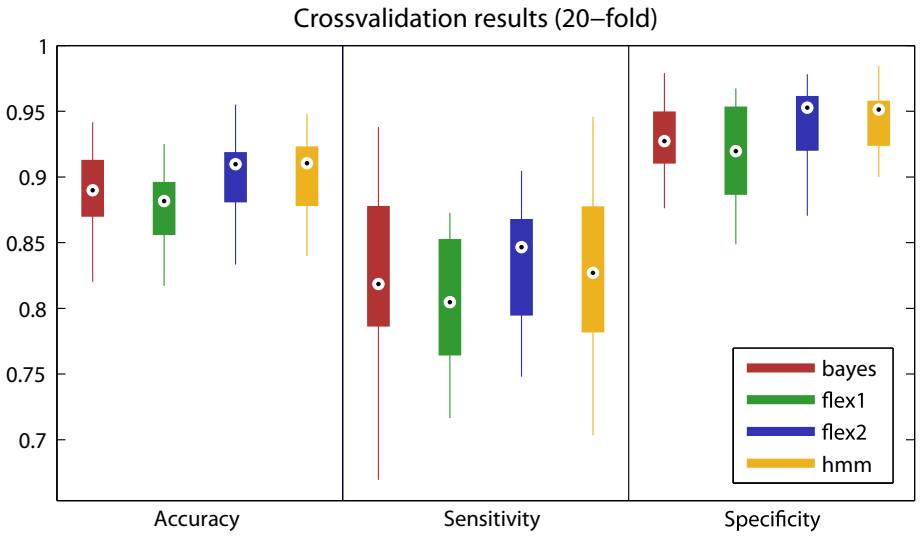

Figure 5.6: Classification results on the 20 validation sets: *bayes* classifier (Moran et al. 2006), Hidden Markov model (*hmm*), suggested model, based exclusively on NRMS (*flex1*) and extended model with added a-priori entry and exit depth distribution (*flex2*).

we can conclude that it is rather the robustness of the NRMS feature itself than the model structure, that has major impact on the results.

The main aim of this chapter was to prove feasibility and efficacy of a probabilistic model which is variable in structure and can potentially be used for fitting of an anatomical 3D model to μEEG signals in multi-electrode setting. In such case, the inside and outside volume of the anatomical model would yield different emission probability distribution and further constraints or penalization on model shift, scaling or rotation could be added easily into the minimization function. We have shown, that such addition of further constituents — such as the entry and exit depth in case of the *flex2* model — can be done and can contribute to improved classification accuracy.

The key part of the presented model is the use of smooth state transition functions, which ensure smooth shape of the resulting likelihood function and enable the use of general-purpose optimization techniques for model fitting. Another consequence of the use of sigmoid transition functions is that the detected transition point does not have to be truncated to a position of available measurement, but can be at an arbitrary position between states (i.e. the detected entry and exit depths are real numbers, not constrained by the depths where μEEG recordings are available).

The drawbacks of the presented model are that contrary to Bayes classifier or an HMM it is not straightforward to convert the presented method to an online algorithm, used e.g. during the surgery. Another weak point is the lack of closed-form solution to model evaluation and

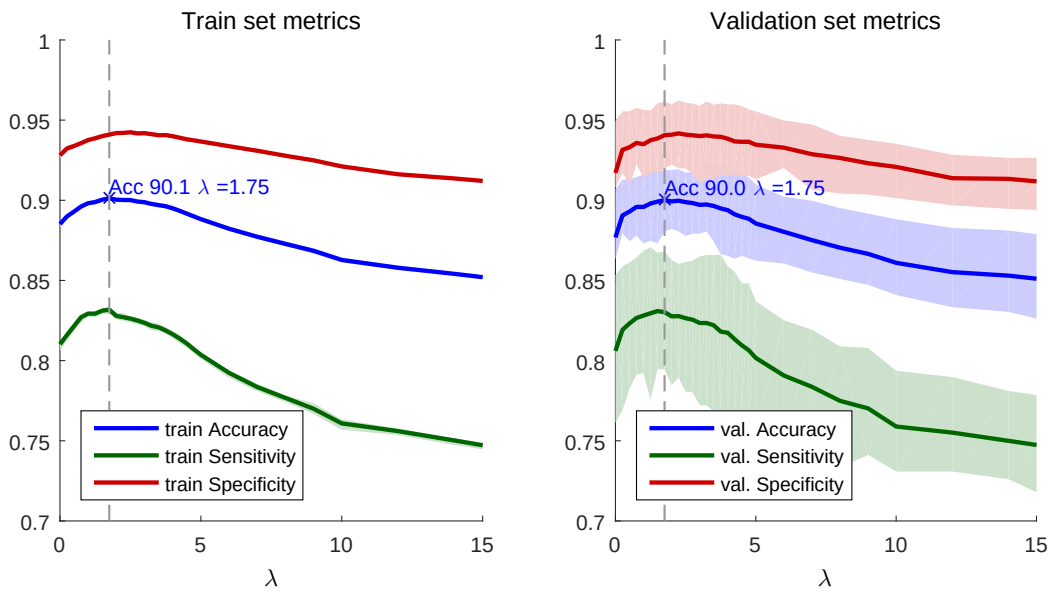

Figure 5.7: Impact of the value of coefficient λ on train (left) and validation set performance. The shaded area represents the inter-quartile range of each metric within values observed on the 20 crossvalidation folds and is therefore comparable to ranges delineated by boxes in Figure 5.6. The variability on the training data is very low due to the high overlap (almost 95%) between training samples.

the necessity to use general optimization. Thanks to the low dimension[6] and small size of the parametric space, this does not pose a real problem in the presented settings, as the parameter estimation took on average 0.9 s on the 247 training trajectories and model evaluation on all 260 trajectories took on average 4.5 s on a standard laptop PC.

5.5 Chapter conclusions

In this chapter, we presented a probabilistic model for identification of the STN in MER exploration trajectories during the DBS surgery. The model uses smooth state transitions and emission probabilities based on observations on a test dataset, consisting of 260 electrode passes in 61 PD patients. Overall, the presented model provided good classification accuracy 90%, which is comparable to other existing solutions based on Hidden Markov Models and Bayes classifier. In the following Chapter (6), the presented model concept serves as a basis for fitting a 3D STN model to the μEEG trajectories directly. This may bring benefits to both target

[6]Dimension of the parametric space searched during the evaluation phase is two, due to two optimized parameters: STN entry a and exit b, both in the range of recorded depths. The search space is further reduced by the conditions defined at the end of Model Evaluation section, especially $a \leq b$.

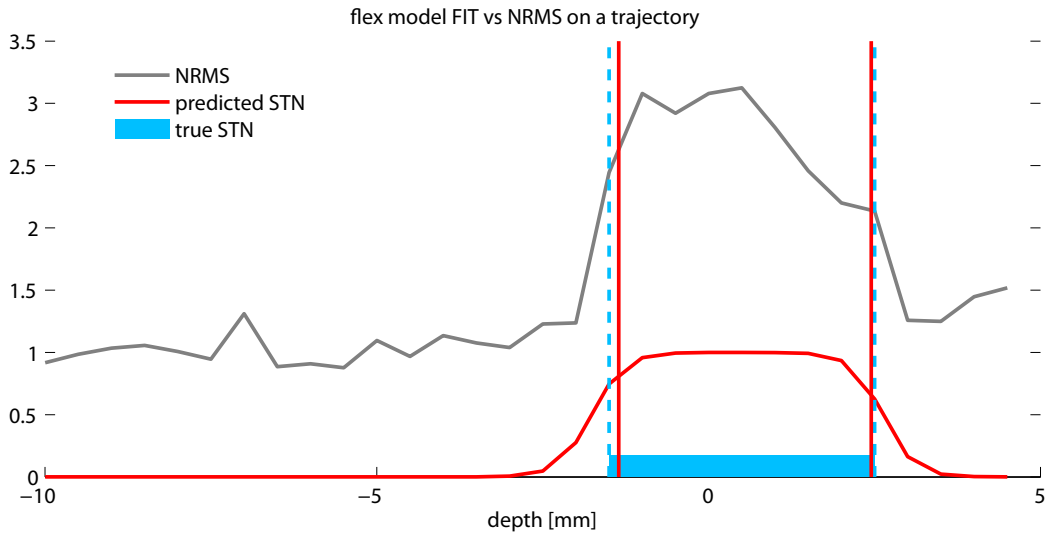

Figure 5.8: Example of *flex1* model fit (red vertical lines — estimated position, red curve — sigmoid weighting function) to a NRMS recorded along a trajectory (grey). The expert-labeled STN position is shown in blue.

identification and modeling of neuronal activity within and around the STN.

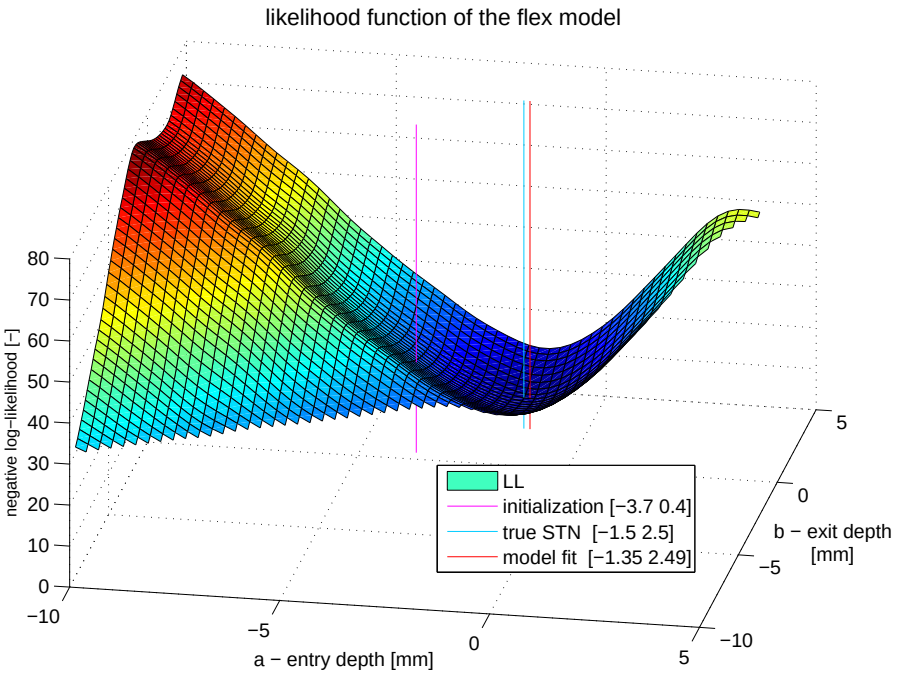

Figure 5.9: Example of the negative log-likelihood function of the *flex1* model shown as a function of hypothetical STN entry (a) and exit (b) depth on a single trajectory. The vertical lines show initialization (magenta), model fit (red) and expert labels (blue).

Chapter 6

Probabilistic fitting of anatomical STN model

Chapter summary

The previous chapter proposed a probabilistic model for classification of MER signals recorded from DBS trajectories and suggested that analogous technique could also be used in the three-dimensional case. In this chapter, we construct such probabilistic model for the fitting of a 3D subthalamic nucleus model to recorded MER signals. We show that due to brain shift and other inaccuracies, occurring during the surgery, the proposed model based solely on recorded electrophysiology provides significantly better classification performance than manual STN identification based on pre-operative MRI imaging. The results suggest that the proposed technique may be used to improve estimation of actual electrode position both intra-operatively and in subsequent data analysis and contribute to higher precision in DBS surgery as well as research of neurophysiological structure of the STN.

6.1 Introduction

As mentioned many times in this thesis, highly accurate electrode placement is necessary in the DBS in order to achieve good symptom suppression and low side effects. This chapter evaluates the possibility of fitting an anatomical 3D model based solely on the measured electrophysiological activity, which might lead to improved localization of the intraoperative MER within the STN and thus lead to more accurate decisions about stimulation contact placement and optimization of the treatment. In order to identify and understand sources of various spatial inaccuracies, the following sections review the standard DBS procedure and suggest which of the shifts can be compensated based on the microexploration MER.

6.1.1 Standard planning procedure

A typical targeting procedure is described in Section 2.2.2 and starts with target identification in pre-surgical imaging and trajectory planning using MRI and possibly also CT scans. In the next stage at the operating theater — the microexploration — a set of microelectrodes is shifted along the planned trajectory, and MER signals are recorded. Based on manual evaluation of the signals by a neurologist, the team decides about stimulation contact placement, removes the microelectrode occupying the chosen trajectory and places the final stimulation electrode. Afterwards, all remaining microelectrodes are also removed and the stimulation electrode is fixed in its place.

Due to brain shifts and electrode bending during the surgery, additional post-operative computed tomography (CT) or MRI imaging may be used to identify accurate final position of the stimulation contact. The post-operative imagery with disctinctive artifact from the stimulation lead and electrode is matched to the pre-operative imagery using transformation based on high-contrast structures (Videen et al. 2008).

Several sources of inaccuracy can be identified in this process, which are summarized and described in the following sections.

6.1.2 Sources of placement error in DBS stereotaxy

Several studies have investigated the various sources of brain shift, occuring between the pre-operative MRI imaging, Microexploration and post-operative CT or MRI and quantified the effect (Nimsky et al. 2000; Pallavaram et al. 2010; Halpern et al. 2008; Ivan et al. 2014). The effects can be summarized as follows:

1. **Brain shift** – Deformation of brain tissue due to pressure changes within the skull (*pneu-*

mocephalus), introduced by the the burr hole[1] and cerebro-spinal fluid leakage during the surgery. The impact of this effect on brain shift between the preoperative MRI and MER was studied and quantified by Pallavaram et al. 2010 – this effect may introduce brain shift of up to 4 mm in the first stage (MRI vs MER), yielding up to 2 mm in the areas near the STN (Martin et al. 2005; Ivan et al. 2014; Pollo et al. 2014).

2. **Electrode bending** – During microelectrode insertion, the straightness of the exploratory path is ensured mainly by high electrode stiffness. The same is true for the insertion of the stimulatory electrode: commonly, the exploration electrode at the selected trajectory is extracted and replaced by the stimulation lead. If present, exploratory electrodes at other trajectories are kept in place to reduce additional shifts. However, electrode bending occurs to some extent in both cases. Lalys et al. have investigated and quantified the extent of stimulation electrode bending based on the post-operative CT (Lalys et al. 2014).

3. **Lead migration** - Lead migration may occur either intra- or post-operatively due to insecure fix of the electrode in the burr hole terminal or patient movement (Bakay et al. 2011). Post-operative electrode bending and migration may also be caused by resolved brain shift, causing discrepancies between post-operative imaging and intra-operative state (Van Den Munckhof et al. 2010).

The extent of brain shift also appeared to affect the number of necessary electrode passes in scenarios where exploration electrodes were added sequentially (Halpern et al. 2008). In other words, higher extent of the brain shift causes a more severe difference between planned and intra-operative state and renders the planned target inaccurate. Some authors suggested that the shift between the pre-operative MRI and intra-operative MER can be reduced by the use of real-time MRI (Martin et al. 2005). However, such equipment is not commonly available in most DBS centers and is not part of the broadly-used procedures (Abosch et al. 2013).

In a research-oriented scenario, where the accurate recording location within the STN is of interest – e.g. to identify internal functional organization (Lourens et al. 2013), location of neuronal populations engaged in a specific function (Sieger et al. 2015), or appropriateness of stimulation contact location (Lalys et al. 2013) – the data are evaluated offline but the requirements for spatial accuracy are very high due to small size of the STN. In such cases, localization of the stimulation electrode contacts in post-operative CT scans is often used to improve localization accuracy. Unfortunately, owing to stimulation electrode bending and migration, the post-operative identification of the electrode contact does not necessarily lead to a more accurate localization of the recording site (Thani et al. 2011).

[1] hole drilled in the skull used to access the brain in stereotactic surgery

6.1.3 Fitting 3D STN models

Several studies have suggested fitting of STN anatomical model to microelectrode recordings. The authors of (Lujan et al. 2009) showed a method to fit a 3D surface atlas to manually classified STN trajectories, based on minimizing the distance of points labeled as STN from the fitted model distance. The authors compare the results to a manual atlas fit to annotated MER points and show superiority of their approach, especially in terms of speed. Similar approach has later been adapted by Lourens et al. 2013 in a study of internal functional organization of the STN. However, the presented solution requires manual inspection and labeling of the MER data and is therefore not fully automatic and does not utilize the full potential of the electrophysiological activity to extract information about STN location in the 3D space. The automatic classification methods, listed in Chapter 5, were restricted to classification of single positions or signle trajectories and do not provide three-dimensional information either.

6.1.4 Available 3D brain atlases

In order to identify surface subthalamic atlases, available for the task of fitting to MER data, we have undertaken a survey of available solutions, summarised in this section.

We sought a high-resolution surface-based model including the STN. Contrary to MRI-based atlases, such as the ATAG atlas (Keuken et al. 2014) derived from 7T MRI imaging or the older BGHAT atlas (Prodoehl et al. 2008) are usually provided in the form of a region of interest (ROI) mask (i.e. voxel positions or probabilities) and are therefore relatively rough ($0.5-1.5\ mm$ voxel size). Other atlases aimed at computing probability maps of the STN to aid identification of optimal stimulation target (Nowinski et al. 2004; Forstmann et al. 2012). However, voxel-based atlases are of a relatively low resolution and are not suitable for the intended model structure.

Other available subcortical atlases that could not be used due to missing or unannotated STN are for instance the Subcortical atlas from the National University of Singapore (Qiu et al. 2010) or the highly detailed atlas by Chakravarty et al. (Chakravarty et al. 2006). The latter was later extended by Sadikot et al. (Sadikot et al. 2011). Another three dimensional atlas of the basal ganglia, including the subthalamic nucleus and substantia nigra was presented by (Yelnik et al. 2007). To the best of my knowledge, this atlas is not publicly available.

Miocinovic et. al provided a three-dimensional stereotactic atlas of the basal ganglia (Miocinovic et al. 2007; Butson et al. 2007), which has been used in similar settings by (Lujan et al. 2009; Lourens et al. 2013; Verhagen et al. 2015). The model used in this thesis is a digitized version of the Anne Morel atlas (Morel 2007) as presented in (Krauth et al. 2010), which is more detailed than that of (Miocinovic et al. 2007) and was kindly provided by prof. Gabor Szekely from VisionLab, ETH, Zurich. The resolution of the mesh model has been reduced by a factor of 5 (987 vertices in the original STN model to 197 in the new model) for performance

reasons. Visualization and comparison of characteristic points before and after the reduction showed no notable differences in model shape or dimensions. The simplified model is visualized in Figure 6.1.

The interested reader can find an extensive and frequently updated summary of available 3D subcortical atlases at the webpage of the Lead-DBS toolbox (*The Lead-DBS project web page*; Horn et al. 2014).

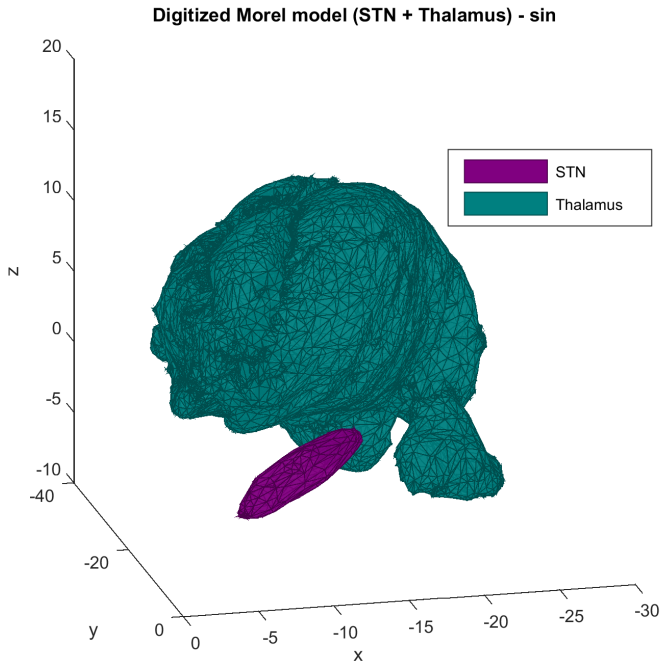

Figure 6.1: Complete 3D Morel atlas (Krauth et al. 2010) showing thalamus (green) and the STN (violet). Only the STN has been used for fitting.

6.1.5 Proposed method

In the presented study, we propose a solution that is based solely on the MER recordings without further requirements on manual signal annotation or cleaning, which could provide information about STN position directly. The model is based on probabilistic modeling of the neuronal background activity, which is higher inside of the STN due to much higher neuron density in the nucleus (see Section 5.2.2 for details on neuronal background activity in the vicinity of and inside the STN). The different activity level inside and outside the nucleus is used to find maximum likelihood fit of a STN surface model (Krauth et al. 2010) and compared to MRI-based model fit in terms of classification accuracy, based on manual expert annotation of the MER. The model itself takes only unprocessed MER signals and information about electrode

direction and planned position as an input and the manual MER labels are used for performance evaluation only.

Contrary to the systems for surgery planning and evaluation such as the Lead-DBS (Horn et al. 2014), the CranialVault (D'Haese et al. 2012) or commercial software provided by producers of DBS solutions such as Medtronic Inc. or AlphaOmega Ltd., aimed mainly at image-based (MRI and/or CT) alignment and surgical planning, the suggested method is aimed on intra- or post-operative identification of location of MER recording sites within the STN. By alignment of MER-based STN fit to the pre-surgical MRI space, the system could be used for evaluation of the extent of the brain shift and help increase precision of the electrode placement. Together with emerging systems aimed at modelling of the stimulation electrical field, such as that presented by Bériault et al. 2012, and emerging directional electrodes (Pollo et al. 2014; McIntyre et al. 2015) the accuracy and efficacy of the DBS procedure may be significantly improved in the upcoming years.

6.2 Methods

The proposed model is based on finding a maximum likelihood fit of a surface STN model to background neuronal activity, extracted from the MER recordings. The model (Krauth et al. 2010) is a 3D surface, represented as a standard triangular 3D mesh, consisting of a set of *vertices* $\boldsymbol{v}$ and a set of *faces* $\boldsymbol{f}$ defining which vertices are connected by surface elements. Translation, scaling and rotation of the model consists of transformation of the matrix of vector points $\boldsymbol{v}$, defined below.

6.2.1 Definition of transformation procedure

We define the 3D transformation used in this study as a matrix operation with nine degrees of freedom, allowing translations t_x, t_y and t_z, scaling factors s_x, s_y and s_z along the x, y and z axis respectively and also rotation around the three axes, given by the angles γ_x, γ_y and γ_z.

The transformation is given by the vector $\boldsymbol{r}$ and can be completely characterized as:

$$\boldsymbol{r} = [t_x, t_y, t_z, s_x, s_y, s_z, \gamma_x, \gamma_y, \gamma_z]. \tag{6.1}$$

The individual components constitute the following matrices:

$$\boldsymbol{R}_x(\gamma_x) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma_x & -\sin\gamma_x \\ 0 & \sin\gamma_x & \cos\gamma_x \end{bmatrix}, \tag{6.2}$$

$$\boldsymbol{R}_y(\gamma_y) = \begin{bmatrix} \cos\gamma_y & 0 & \sin\gamma_y \\ 0 & 1 & 0 \\ -\sin\gamma_y & 0 & \cos\gamma_y \end{bmatrix}, \tag{6.3}$$

$$\boldsymbol{R}_z(\gamma_z) = \begin{bmatrix} \cos\gamma_z & -\sin\gamma_z & 0 \\ \sin\gamma_z & \cos\gamma_z & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6.4}$$

for rotation. Model scaling is given by the matrix $\boldsymbol{S}$

$$\boldsymbol{S} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & s_z \end{bmatrix} \tag{6.5}$$

$$\tag{6.6}$$

and translation by the vector $\boldsymbol{t}$:

$$\boldsymbol{t} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \tag{6.7}$$

The transform of a column vector $\boldsymbol{x}$ using parameters $\boldsymbol{r}$ can then be expressed as:

$$\boldsymbol{x}' = T(\boldsymbol{x}, \boldsymbol{r}) = \boldsymbol{S}\boldsymbol{R}_x\boldsymbol{R}_y\boldsymbol{R}_z\boldsymbol{x} + \boldsymbol{t}, \tag{6.8}$$

using the definition of $\boldsymbol{R}, \boldsymbol{S}, \boldsymbol{t}$ above. In most cases, more than one point needs to be transformed. In such case, the formula from 6.8 can be used analogously, with data matrix $\boldsymbol{X}$, composed of p column vectors $[\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_p]$:

$$\boldsymbol{X} = \begin{bmatrix} x_{1,1} & x_{2,1} & \dots & x_{p,1} \\ x_{1,2} & x_{2,2} & \dots & x_{p,2} \\ x_{1,3} & x_{2,3} & \dots & x_{p,3} \end{bmatrix}, \tag{6.9}$$

where $x_{i,j}$ is j-th element of i-th point. The resulting transformation using parameter vector $\boldsymbol{r}$ is defined as:

$$\boldsymbol{X}' = T(\boldsymbol{X}, \boldsymbol{r}) = \boldsymbol{S}\boldsymbol{R}_x\boldsymbol{R}_y\boldsymbol{R}_z\boldsymbol{X} + \boldsymbol{t}\boldsymbol{J}_{1,p} \tag{6.10}$$

where $\boldsymbol{X}'$ is the matrix of transformed column vectors $[\boldsymbol{x}'_1, \boldsymbol{x}'_2, ..., \boldsymbol{x}'_p]$ and $\boldsymbol{J}_{1,p}$ is an all-ones matrix with one row and p columns.

The coordinate system used for all visualizations is shown in Figure 6.2 and comprises x axis pointing from the left to the right hemisphere (same for both hemispheres, meaning lateral to medial direction in case of the left hemisphere), y axis pointing posterior to anterior direction and z axis pointing ventral to dorsal.

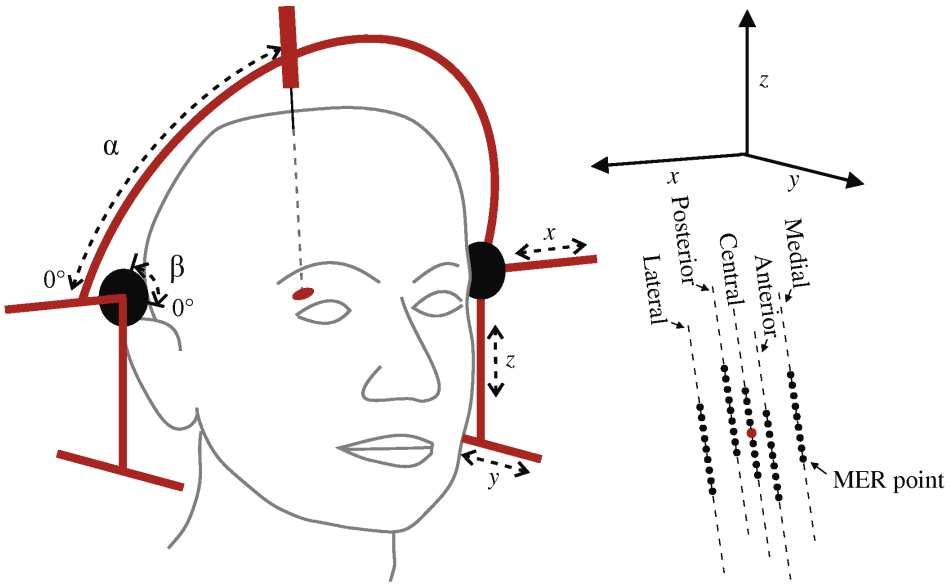

Figure 6.2: Coordinate system, used in the study, showing directions of the main axes and angles α and β on the stereotactic frame. Image adapted from Lourens et al. 2013.

6.2.2 Model structure

The basic model structure is similar to the one-dimensional case presented in Chapter 5, Section 5.2.3: First, neuronal background activity level is estimated from each MER signal using the NRMS measure proposed by Moran et al. 2006 and described in detail in Section 5.2.2. Next, the expected distribution of the NRMS values in each state is modeled by the log-normal distribution.

Contrary to the aforementioned single-trajectory model where the state sequence of "pre-STN, STN, post-STN" can be assumed, the three-dimensional case requires modified structure, as the recording site may be in arbitrary position with respect to the 3D model. We designed the model with only two states: i) Inside the STN (IN) and ii) outside the STN (OUT), with a single state membership logistic function, defining the smooth transition between the two states. The general idea of the model remains the same and is based on finding a transformation vector $\boldsymbol{r}^*$ which optimizes the likelihood of producing a set of observations (i.e. NRMS values)

$\boldsymbol{x} = \{x_1, ..., x_N\}$ recorded at locations $\boldsymbol{L} = \{\boldsymbol{l_1}, ..., \boldsymbol{l_N}\}$, where $\boldsymbol{l_i}$ are the 3D recording site coordinates corresponding to observation x_i. The transformation $T(\boldsymbol{v}, \boldsymbol{r^*})$, according to Eq. 6.8 is applied to the vector $\boldsymbol{v}$ of model vertices at the initial position.

The following paragraph define the probabilistic model components and explain computation of the likelihood function

Emission probabilities

Similarly to the 1D version of the model, the emission probabilities represent how likely a background activity (NRMS) level x_i is to be observed in the respective state. The emission probabilities are modeled using the log-normal distribution, as depicted in Figure 6.3. Parameters of the log-normal distribution for both states are estimated in the training phase on training data using the standard maximum likelihood estimates, already defined in Eq. 5.1 and 5.2, the only difference being joining the states *pre* and *post* into a single state. As a result, the two parameters are estimated for each state, leading to estimated parameter values $\{\hat{\mu}_{OUT}, \hat{\sigma}_{OUT}, \hat{\mu}_{IN}, \hat{\sigma}_{IN}\}$, which become part of the parameter vector $\boldsymbol{\Theta}$.

In the evaluation phase, the emission probability $p(x_i|s, \boldsymbol{\Theta})$ of observing NRMS value x_i in a state s given model parameters $\boldsymbol{\Theta}$ is calculated using formula for probability density function of the log-normal distribution, according to:

$$p(x_i|s, \boldsymbol{\Theta}) = \frac{1}{x_i \hat{\sigma}_s \sqrt{2\pi}} \quad \exp - \frac{(\ln(x_i) - \hat{\mu}_s)^2}{2\hat{\sigma}_s^2}, \tag{6.11}$$

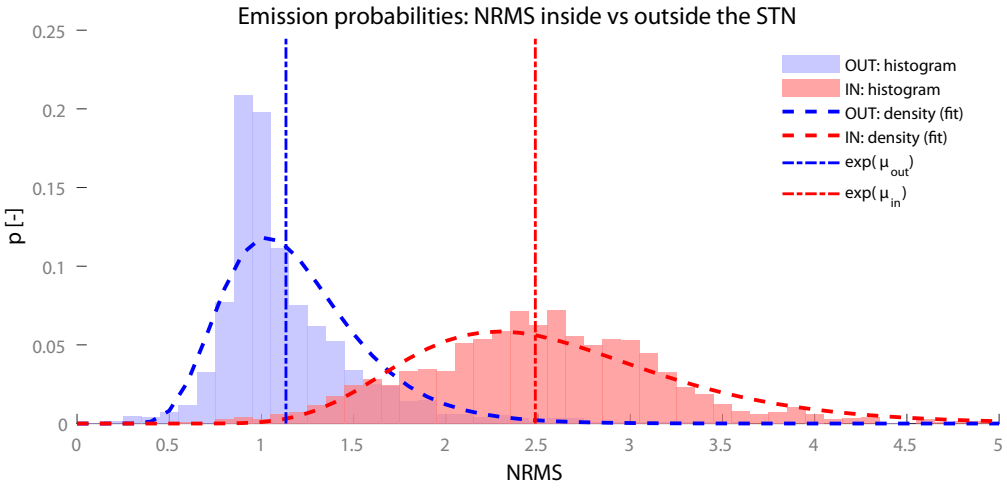

Figure 6.3: Fitted emission probabilities: histograms of observed NRMS values inside (red area) and outside (blue area) the STN, with fitted log-normal probability density functions (dashed curves) and their parameters (vertical lines).

Membership probabilities

In order to retain the smooth gradient function, discussed in the previous Chapter, the 3D model also includes the assumption of a smooth state transition, modeled by logistic sigmoid function. As there are only two states in this model, a single sigmoid function S is sufficient to maintain the transition (as opposed to two sigmoid functions S_{en} and S_{ex} in case of the 1D model). As the slope of the transition is different near STN entry and STN exit and only data along the electrode trajectories are available at sufficient resolution (only very rough data would be available to estimate parameters of the transition in other directions from the STN), the single sigmoid S is trained on training nrms data aligned with respect to the STN entry, combined with mirrored data aligned with respect to the STN exit. The situation is depicted in Figure 6.4, where the entry data are shown in blue and exit data in green and the joint sigmoid S (red) can be compared to separate S_{en} (blue) and S_{ex} (green) sigmoids, fitted to the entry and exit data separately.

During *training*, an extended function S' with one extra degree of freedom is used to find optimum parameters:

$$S'(d'_i) = 1 + \alpha^1 \cdot \left(1 + \exp -(\beta^0 + \beta^1(d_i))\right)^{-1}, \tag{6.12}$$

with d'_i being the distance to the transition (to STN entry or negative to STN exit) and α^1 being the auxiliary scaling variable, replaced by the emission probabilities later during the evaluation phase.

Values of the parameters $\{\alpha^1, \beta^0, \beta^1\}$ are found as to minimize mean square error on training data:

$$\underset{\alpha^1,\beta^0,\beta^1}{arg\,\min} \quad \sum \sum_{i=1}^{M} \left(S'(d'_i, \alpha^1, \beta^0, \beta^1) - x_i\right)^2, \tag{6.13}$$

where x_i is the NRMS value measured at a distance d'_i from the STN boundary for M training samples.

In the *evaluation phase*, the sigmoid transition function depends only on distance d_i from model surface, rotated using vector $\boldsymbol{r}$ and is computed according to:

$$S(\boldsymbol{d_i}|\boldsymbol{\Theta}) = \left(1 + \exp -(\beta^0 + \beta^1(d_i))\right)^{-1} \tag{6.14}$$

It is important to note that d_i is computed as euclidean distance between the measurement location $\boldsymbol{l_i}$ and the nearest surface point of the STN model at current location and may be formalized as $d_i = d_{surf}(\boldsymbol{l_i}, \boldsymbol{r}) = \|\boldsymbol{l_i} - \boldsymbol{a_i}\|$, where $\boldsymbol{a_i}$ is the nearest surface point to $\boldsymbol{l_i}$ on the STN model transformed by $\boldsymbol{r}$. For the sake of simplicity, we will assume the relation between $\boldsymbol{l_i}$ and d_i as implicit in the following. Additionally, the distance is multiplied by -1 if the location $\boldsymbol{l_i}$ lies outside of the model and by +1 when inside.

The membership probabilities for trained model parameters $\boldsymbol{\Theta}$ and anatomical model transformed by the vector $\boldsymbol{r}$ are then computed according to:

$$p(\boldsymbol{l_i} \in IN|\boldsymbol{r}, \boldsymbol{\Theta}) = S(l_i|\boldsymbol{r}, \boldsymbol{\Theta}) \tag{6.15}$$

for the state IN and:

$$p(\boldsymbol{l_i} \in OUT|\boldsymbol{r}, \boldsymbol{\Theta}) = 1 - p(\boldsymbol{l_i} \in IN|\boldsymbol{r}, \boldsymbol{\Theta}) = 1 - S(l_i|\boldsymbol{r}, \boldsymbol{\Theta}) \tag{6.16}$$

for the state OUT. Note that thanks to the model consisting only of two states, the membership probabilities add to one implicitly and no normalization constant, simiar to z_i in the 1D model is necessary.

The trained model is fully characterized by the parameter vector $\boldsymbol{\Theta} = \{\hat{\mu}_{OUT}, \hat{\sigma}_{OUT}, \hat{\mu}_{IN}, \hat{\sigma}_{IN}, \beta^0, \beta^1\}$, comprising parameters of the emission probability densities and those of the sigmoid function.

Likelihood function and MLE estimation

The joint probability of being in state s and observing a NRMS value x_i at position $\boldsymbol{l_i}$ is computed from the emission and membership probability functions according to the Bayes' theorem:

$$p(\{x_i, \boldsymbol{l_i} \in s\}|\boldsymbol{r}, \boldsymbol{\Theta}) = p(x_i|\boldsymbol{l_i} \in s, \boldsymbol{r}, \boldsymbol{\Theta}) \cdot p(\boldsymbol{l_i} \in s|\boldsymbol{r}, \boldsymbol{\Theta}) \tag{6.17}$$

The likelihood function for a single observation is then computed as a marginalization over both states (IN, OUT):

$$\begin{aligned} L(\{x_i, \boldsymbol{l_i}\}|\boldsymbol{r}, \boldsymbol{\Theta}) &= p(\{x_i, \boldsymbol{l_i}\}|\boldsymbol{r}, \boldsymbol{\Theta}) = \\ &= p(\{x_i, \boldsymbol{l_i} \in IN\}|\boldsymbol{r}, \boldsymbol{\Theta}) \\ &\quad + p(\{x_i, \boldsymbol{l_i} \in OUT\}|\boldsymbol{r}, \boldsymbol{\Theta}) \end{aligned} \tag{6.18}$$

And by using the expansion according to Eq. 6.17, the full likelihood for a single observation yields:

$$\begin{aligned} L(\{x_i, \boldsymbol{l_i}\}|\boldsymbol{r}, \boldsymbol{\Theta}) &= p(\{x_i, \boldsymbol{l_i}\}|\boldsymbol{r}, \boldsymbol{\Theta}) = \\ &= p(x_i|\boldsymbol{l_i} \in IN, \boldsymbol{r}, \boldsymbol{\Theta}) \cdot p(\boldsymbol{l_i} \in IN|\boldsymbol{r}, \boldsymbol{\Theta}) \\ &\quad + p(x_i|\boldsymbol{l_i} \in OUT, \boldsymbol{r}, \boldsymbol{\Theta}) \cdot p(\boldsymbol{l_i} \in OUT|\boldsymbol{r}, \boldsymbol{\Theta}). \end{aligned} \tag{6.19}$$

To compute the joint likelihood of the whole observation sequence $\boldsymbol{x} = \{x_1, ..., x_N\}$, $\boldsymbol{L} = \{\boldsymbol{l_1}, ..., \boldsymbol{l_N}\}$, we naïvely assume conditional independence given model parameters and compute the joint likelihood as:

$$L(\{\boldsymbol{x}, \boldsymbol{L}\}|\boldsymbol{r}, \boldsymbol{\Theta}) = \prod_{i=1}^{N} L(\{x_i, \boldsymbol{l_i}\}|\boldsymbol{r}, \boldsymbol{\Theta}) \tag{6.20}$$

Similarly to the case of the 1D model (Eq 5.8), the optimization procedure aims to find a model transformation given by the vector $\boldsymbol{r}$, which maximizes the likelihood of the observed data. For numerical stability, we search for the minimum of the negative log-likelihood function, which is equivalent to the maximum of the original likelihood function owing to monotonicity of the $ln()$ operation. The minimization then has the form:

$$\boldsymbol{r}^* = arg\min_{\boldsymbol{r}} \sum_{i=1}^{N} -\ln(L(\{x_i, \boldsymbol{l_i}\}|\boldsymbol{r}, \boldsymbol{\Theta})), \tag{6.21}$$

where $\boldsymbol{r}^*$ is the MLE estimate of optimal transformation parameters, given the model parameters and the observation sequence. The minimization is performed using general purpose constrained optimization (the *active set* algorithm as implemented in MathWorks Matlab `fmincon` function). Initial model position is defined by the *target* method described below and the initial transformation vector is $r = [0, 0, 0, 1, 1, 1, 0, 0, 0]$, i.e. no rotation and translation and identity scaling in all dimensions. To prevent the model from diverging from clinically reasonable scaling and rotation, we set the maximum shift to ± 5 mm in any direction, maximum scaling $\pm 25\%$ in each direction and rotation maximum $\pm 15°$ around each axis.

6.2.3 Reference methods

In order to evaluate performance of the proposed method, we implemented two reference methods, based solely on anatomical landmarks, identified manually by neurologists in the pre-operative MRI images:

1. **target** – the method consists in finding a translation $[t_x, t_y, t_z]$, which shifts central point of the atlas model to the planned target point. No rotation or scaling are considered and the translation can thus be computed as a difference between the two points. The *target* method is also used as the initalization for NRMS-based fitting, as it requires no additional information apart from planned target coordinates, which is the result of standard pre-surgical planning procedure.

2. **acpc** – this method represents a simple atlas fitting approach, based on two significant brain landmarks: the anterior commisure (AC) and the posterior commisure (PC), which are clearly visible in the MRI scans and are often used as a reference for spatial localization of deep brain structures. The method analytically finds a full transformation vector $\hat{\boldsymbol{r}}$ (i.e. including translation, scaling and rotation along all axes), which maps the vector given by AC and PC points in the atlas to the vector given by AC-PC points, identified in patient's MRI scans.

3. **allpoints** – additionally to the AC-PC points, this method uses 12 landmarks on the

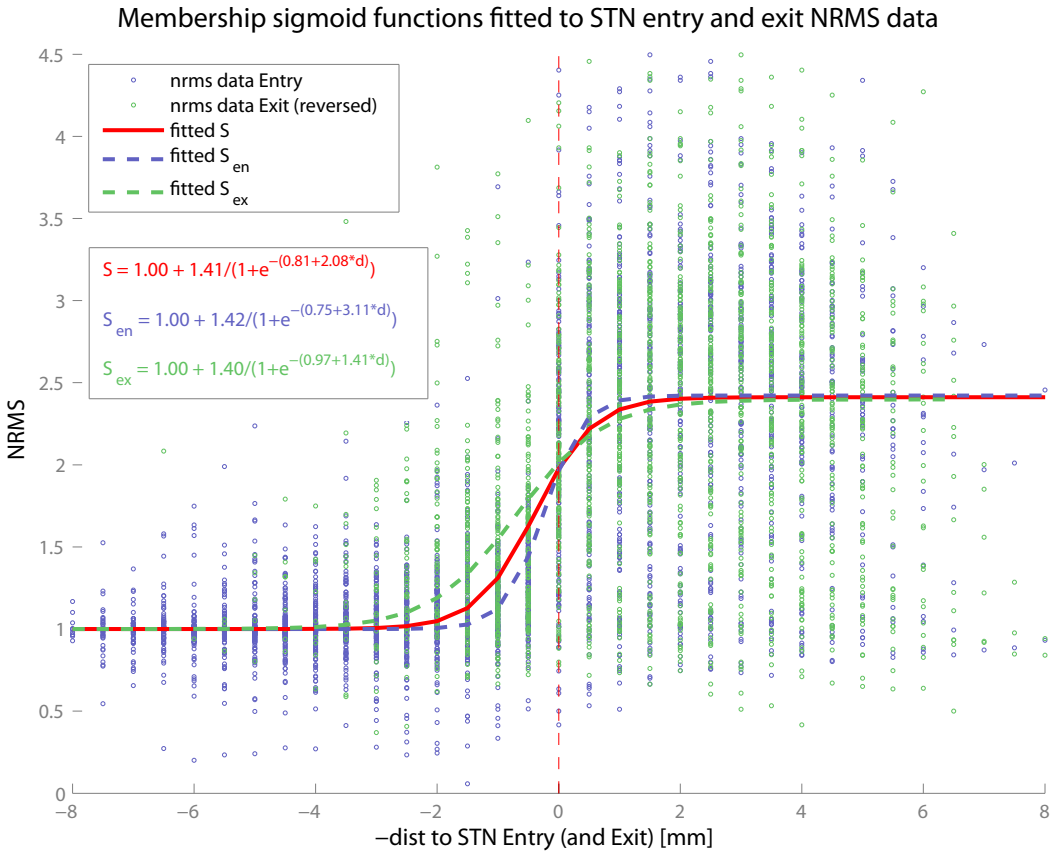

Figure 6.4: The membership logistic sigmoid function S (red) fitted to measured NRMS data around the STN entry (blue circles) and exit (green circles, depth-flipped/negative) data. The fitted sigmoid S can be compared to separate entry and exit sigmoid S_{en} and S_{ex} from the previous chapter (dashed blue and green curves). Compare to Figure 5.4: different values of the two fitted sigmoids are caused by different (smaller) data set including only patients with 3D reference information.

STN boundaries, defined previously in (Sieger et al. 2015). The exact definition of the landmark points is given in Table 6.1. Based on these definitions, the landmark points were transferred analytically to the STN model, which is visualized in Figure 6.5.

Finding the transformation parameters $\hat{r}$ which fit the pairs of 14 atlas points $P_i^m \in \{P_1^m, ..., P_1^m 4\}$ to the points identified in patient MRI scans $P_i^p \in \{P_1^p, ..., P_n^p\}$ is an overdetermined task, which is solved by the least-squares estimation according to:

$$\hat{\boldsymbol{r}} = \underset{\boldsymbol{r}}{arg\,min} \quad \sum_{i=1}^{n} \|P_i^p - T(P_i^m, \boldsymbol{r})\| \tag{6.22}$$

where the $\|\cdot\|$ represents the l^2 norm, $T()$ is transformation operation from Eq. 6.8 and $n = 14$ and includes the 12 landmark points, the AC and PC.

Table 6.1: STN delineating points: definition of key points as defined in (Sieger et al. 2015). Visualization can be found in Fig. 6.5

Point	location definition
P_1	Most ventral point in the posterior boundary
P_2	Most dorsal point in the anterior boundary
P_3	Most medial point in the in the most medially protruded slice in the anterior part of the STN
P_4	Most lateral point in the most medially protruded slice in the anterior part of the STN
P_5	Medial intersection of the axial plane crossing the middle point and the middle contour
P_6	Lateral intersection of the of the axial plane crossing the middle point and the middle contour
P_7	Dorsal intersection of the sagittal plane crossing the middle point and the middle contour
P_8	Ventral intersection of the sagittal plane crossing the middle point and the middle contour
P_9	Most medial point on the middle contour
P_{10}	Most lateral point on the middle contour
P_{11}	Most dorsal point on the middle contour
P_{12}	Most ventral point on the middle contour

The middle point is the center of the line connecting points P1 and P2. The middle contour is the coronal contour intersecting the middle point.

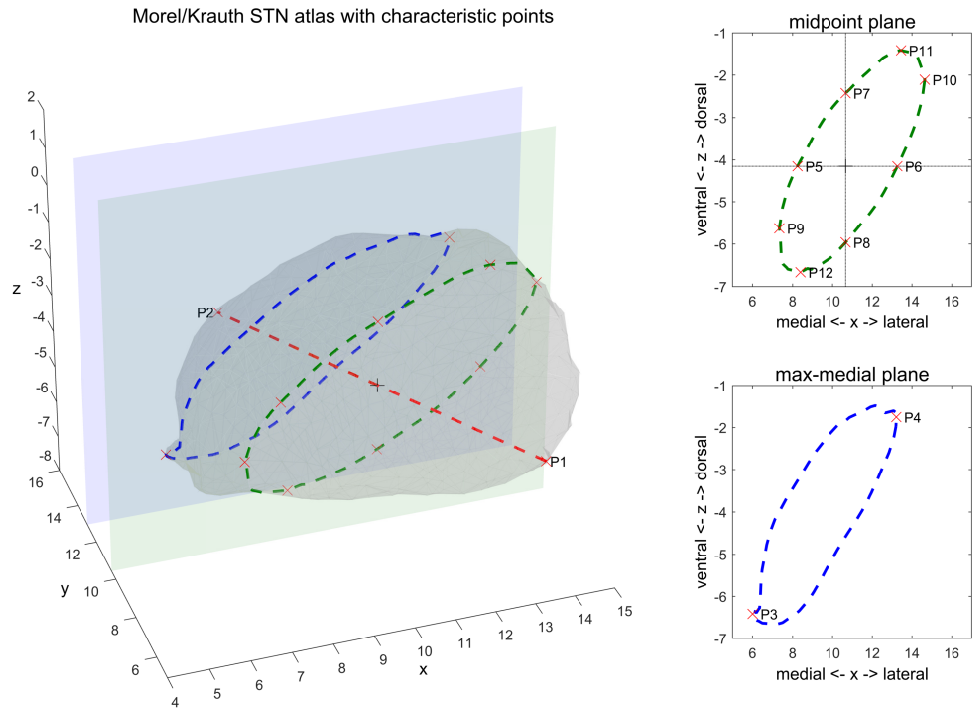

Figure 6.5: Illustration of characteristic points as defined and used previously by our research group in (Sieger et al. 2015), transferred newly to the 3D Morel atlas (Krauth et al. 2010) according to their definition in Table 6.1

6.2.4 Data collection and preprocessing

The MER signals were recorded intra-operatively from five parallel electrode trajectories, spaced 2mm apart in a cruciform configuration around the central electrode. The spatial configuration can be seen in Figure 6.2. The sampling frequency was 24 kHz, signals were filtered by a bandpass filter in the range 500–5000 Hz upon recording and stored for offline processing. At each of the recording positions, spaced apart by 0.5 mm, a typically ten seconds of MER signal were recorded using each electrode.

During preprocessing, the most distinct artifact have been detected using the extended *COV* method presented in Chapter 4. The reason not to use one of the more accurate methods was that due to the simple feature calculation, only the most pronounced frequency artifacts can introduce noise to the results. As such, the COV method was chosen to retain at least some portion of signal at each depth (to avoid excessive amounts of missing data) at the cost of noise in the computed neuronal background levels. An example of MER signals measured at a set of five trajectories can be found in Figure 6.6.

For the purpose of model fit evaluation, manual intra-operative expert annotation of the MER signals has been stored, assigning each signal to a nucleus of origin: the *STN*, *SNr* (substantia nigra), *thalamus* or other. In the evaluation, the STN was used as the target class

(corresponding to the state *IN*), while all other states were joint under *other* (corresponding to the state *OUT*).

To compute the reference fitting methods, based on anatomical landmarks, the center of the anterior commisure (AC), center of the posterior commisure (PC) and the 12 landmark points listed in Table 6.1 (see also Figure 6.5) were identified in the pre-operative MRI images by a trained neurologist.

Additionally to the landmark points and nuclei annotations, coordinates of the planned target, as well as setting of the stereotactic frame, defined by the anlges α and β has been collected.

The background neuronal activity level was computed using the normalized root-mean-square (NRMS) according to (Moran et al. 2006) in a process described above in 5.2.2.

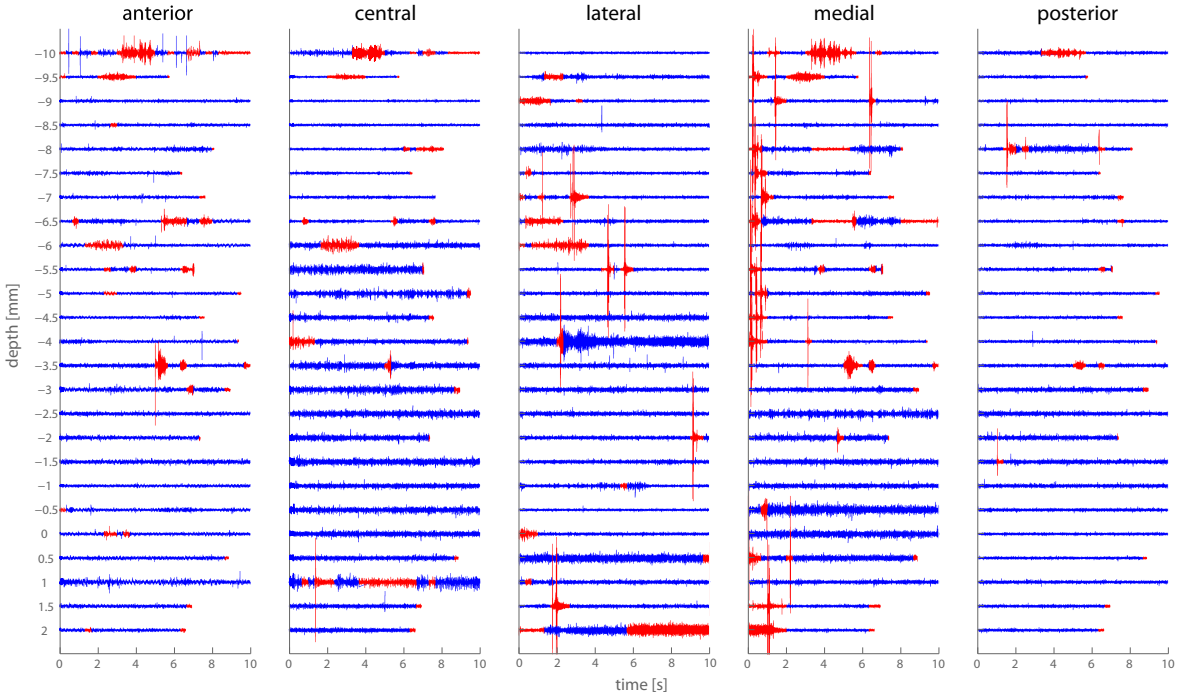

Figure 6.6: Example of one DBS MER exploration with five parallel electrode trajectories. Artifacts detected using the extended COV method presented in Chapter 4 are shown in red.

6.2.5 Performance evaluation

To estimate the out of sample performance of the proposed method and due to the relatively small sample size (in terms of whole patient sets), we employed the following leave one subject out (LOSO) procedure:

1. In each iteration, data from a single patient (yielding one or two explorations) were chosen as the evaluation sample and kept aside.
2. The probabilistic model parameters $\boldsymbol{\Theta}$ were computed on the remaining part of the dataset
3. Each of the the explorations from the validation set (maximum two in a bi-laterally implanted patient) was used to find the NRMS-based maximum-likelihood model fit $\boldsymbol{r}^*$. Additionally the three anatomical reference methods were performed on each of the validation trajectories and all results were stored.
4. The process was repeated until all patients' data were used for validation and the results were aggregated, taking sample size of each patient into account.

To evaluate quality of the model fit, we used the machine-learning based approach used also by Lujan et al. 2009. We calculated the number of recording positions, correctly classified by the resulting model as follows:

- TP – true positives – recordings labeled by the expert as coming from the STN at locations encapsulated within the surface of the trasformed model (i.e. "inside the model")
- FP – false positives – non-STN points (from all *other* nuclei), inside of the model
- TN – true negatives – non-STN points outside of the transformed model
- FN – false negatives – STN points outside of the model

Using these definitions, the accuracy, sensitivity, specificity and Youden J-index ($J = sens + spec - 1$) were computed according to standard definition, with STN as the positive class.

6.3 Results and discussion

This section summarizes and comments on the results of performance evaluation on the collected database of MER signals.

6.3.1 Collected data

The dataset contained data from 27 explorations in 15 PD patients with complete 3D information and additional 8 explorations from 4 patients with measured and annotated MER signals but without information on planned target position and stereotactic frame settings. The latter small set was included for estimation of model parameters ($\boldsymbol{\Theta}$) but excluded from validation. Each exploration consisted of 5 electrode trajectories with 25.9 recording positions on average.

Table 6.2: Overall 3D STN model fitting crossvalidation results on the 27 validation trajectories for all methods.

	Accuracy		Sensitivity		Specificity		Youden J	
method	**mean**	std	**mean**	std	**mean**	std	**mean**	std
target	**74,3%**	7,6%	**40,7%**	12,3%	**87,3%**	5,7%	**28,0%**	17,0%
acpc	**75,7%**	8,9%	**44,7%**	17,2%	**87,6%**	8,2%	**32,3%**	21,1%
allPoints	**78,7%**	8,7%	**44,6%**	19,8%	**92,3%**	4,9%	**36,8%**	21,3%
nrmsCon	**88,1%**	5,2%	**69,0%**	14,2%	**95,5%**	5,4%	**64,5%**	13,6%

In total, the data included 35 explorations from 19 patients, leading to 175 electrode trajectories and 4538 recorded MER signals.

6.3.2 Performance evaluation

Classification performance (the proportion of correctly included/excluded STN points) was evaluated for each of the fitting methods on the 27 exploration trajectories, the results are shown in the Figure 6.7 and Table 6.2. As seen from both classification representations, it is clear that the nrms-based method provided substantially better fit to the measured MER sites than any of the other method. The results further show, that the main difference is driven especially by the higher sensitivity, i.e. the proportion of correctly included STN points inside the model. This is even more clearly seen from the tabulated values of the Youden J statistic, where the proposed method surpasses the reference methods by a factor of two. It has to be considered that the dataset is highly imbalanced dataset with only 27% of signals coming from the STN.

The classification accuracy represents only one possible view on goodness of model fit. Therefore, the following section evaluates the actual transformation parameters.

6.3.3 Analysis of transformation parameters

Additional important information may be provided by analysis of transformation parameters of the fitted models. We may assume, that the landmark-based *allPoints* method will provide a very good fit in terms of model scaling and rotation, which are not as affected by the brain shift as the translation parameters. We would therefore expect the *nrmsCon* model to provide similar results in terms of rotation and scaling and highly different translation. The comparison of model parameters of the two models is shown in the Figure 6.8.

As seen from the figure, the difference in translation t_y along the y (antero-posterior anatomi-

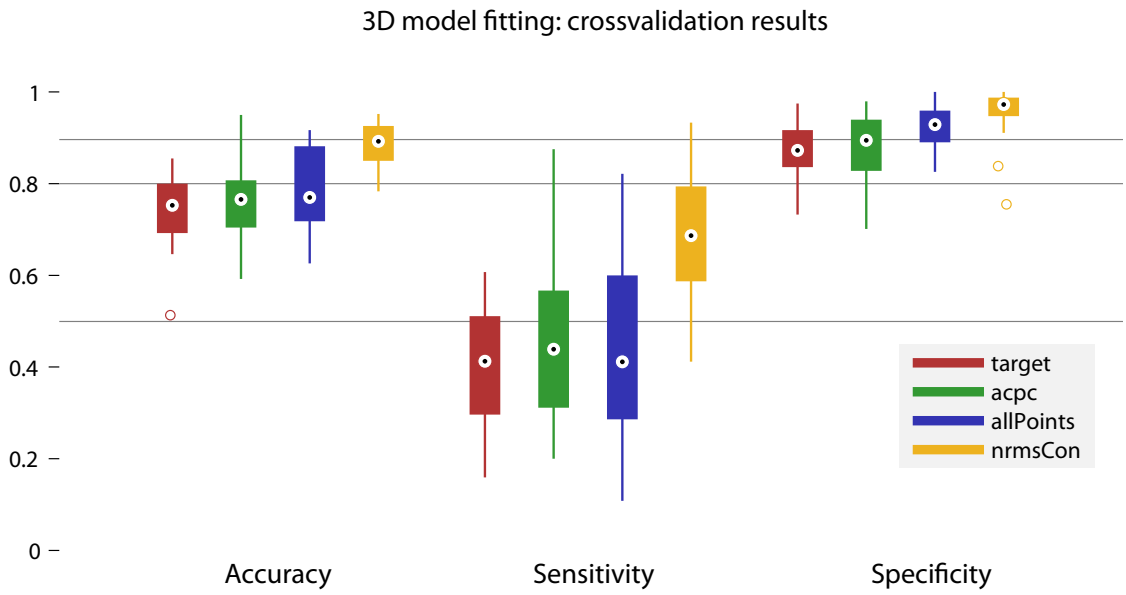

Figure 6.7: Comparison of different fitting methods in terms of classification performance (correctly included/excluded recorded NRMS points): the electrophysology-based proposed method *nrms* (yellow) showed higher STN identification accuracy than the reference MRI-based methods, especially the *allPoints* (blue) based on fitting of 12 landmark points and the AC-PC. The performance advantage lays mainly in much improved sensitivity, compared to the rest of the methods

cal direction) is apparent, while the differences in translation along other axes are not so striking. According to previous studies (Ivan et al. 2014; Pallavaram et al. 2010), the brain shift occurs predominantly in the posterior direction (negative shift along the y axis) due to gravity and patient being operated in lying position. Considering the much better fit to the measured electrophysiology locations, achieved by the *nrmsCon* method, we may assumed that the proposed method managed to correctly identify shift of the brain issue with respect to pre-operative MRI scans.

Evaluating the scaling and rotation parameters, there is clearly a higher degree of variability in the case of the *nrmsCon* model, which may signalize a slightly higher anatomical variability in model positioning than in case of the *allPoints* method. However, the constraints, defined in Section 6.2.2 limit the transformation in a desired range and can be changed easily.

6.3.4 Discussion

The proposed *nrmsCon* method achieved highly increased performance in STN identification in terms of Accuracy, sensitivity and specificity compared to the reference MER landmark-based *allPoints*. Compared to previous studies on automatic MER classification, which aim to assist

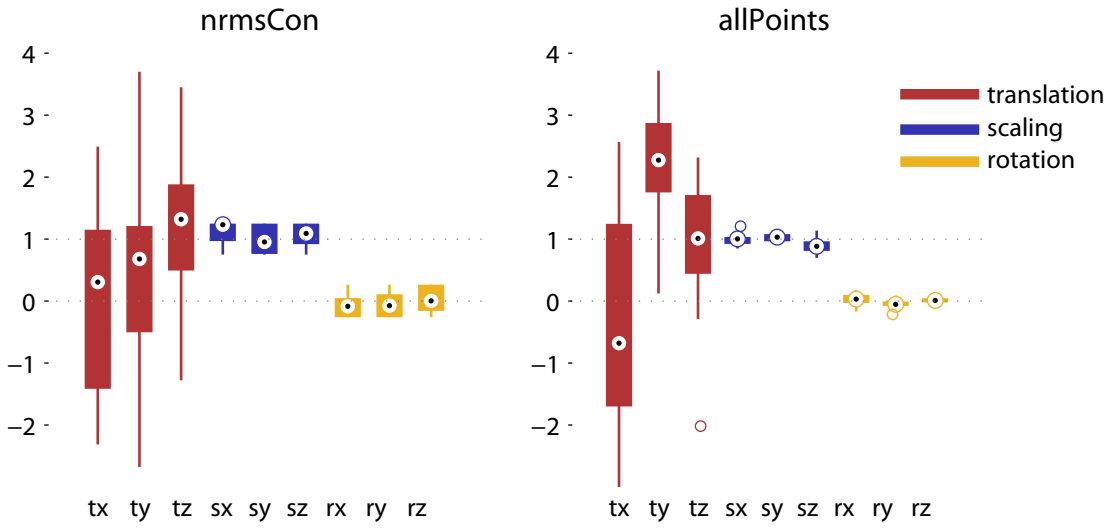

Figure 6.8: Boxplot of the transformation parameters computed by the proposed method method *nrmsCon*, compared to the reference method *allPoints*, based on 14 MRI landmarks. The parameters include translations t (red), scaling s (blue) and rotation r (yellow, in radians). Note the significant difference in t_y - the main direction of the brain shift. The apparently non-zero shift in t_z suggests the actual target in pre-surgical planning may be slightly more ventrally placed compared to the atlas-delineated center point.

or replace the manual MER labelling during the surgery (Reviewed in Chapter 5), the proposed model introduces the benefit of anatomical information, providing a means to evaluate extent of the brain shift and accurate positioning of recording site within an anatomical model. As opposed to previous studies using anatomical model fitting (Lujan et al. 2009; Lourens et al. 2013), the proposed method is fully automatic and requires no manual labeling or other prior knowledge other than location of the planned target and angular settings of the stereotactic frame, which is known prior to the surgery.

It may be objected, that this increase is achieved at the cost of a less anatomically accurate model position – in other words model overfitting to the data. Fortunately, several options exist to achieve balance between the accuracy of MER classification and a trustworthy model positioning:

1. Instead of using hard limits on various parameters of the rotation vector r, the likelihood may be extended with a-priori distribution of each parameter values. This has already been suggested and verified on the 1D model (*flex2*) in the previous chapter. The a-priori distribution of transformation parameters may be based either on published studies (such as Daniluk et al. 2010; Keuken et al. 2014, especially for shift) or from distribution of models fitted using the *allPoints* method (especially for scaling and rotation, which may

be assumed more robust against shift-induced error).

2. For applications where high accuracy is of interest – such as in the case of research studies – the both approaches may be combined: based on the MRI landmarks, optimal model rotation and scaling may be computed using the *allPoints* method. Subsequently, these parameters may be fixed and optimal translation may be computed using the *nrmsCon* method.

In the proposed scenario, we searched optimal fitting of an anatomical atlas to the recorded MER signals, which is useful especially for identification of particular region within the STN on a common atlas model. As suggested above, the system could be used instead for mapping of the intra-operative space, affected by brain shift to the pre-surgical MRI scans. This way, improved visualization and target localization might be obtained.

In terms of computational time, the proposed method required on average 29.3 s (std 10.1 s) to fit the model to patient data on a standard laptop PC. It may be assumed, that code optimization might lead to significant improvement in this already acceptable parameter, especially considering no manual data labeling or preprocessing is necessary.

Overall, our study showed high potential of the presented probabilistic model in identification of accurate target position and we find the results very promising for future research in electrophysiology-based atlas mapping.

6.4 Chapter summary

In this Chapter, we have proposed a novel method to fit a three-dimensional STN model based solely on the electrophysiological activity, measured along exploration trajectories. Compared to the reference method *allPoints*, based on 14 landmark points identified in the pre-operative MRI scans, the proposed method achieved dramatically increased performance in identifying the STN nucleus in all measures: mean accuracy was 88.1% as opposed to 78.7% in the case of the *allPoints method*, while the advantage in sensitivity was even more pronounced: 69.0% vs 44.6%, maintaining also higher specificity.

If appropriately applied, the proposed method may increase accuracy to the surgical procedure by accurately evaluating the extent of electrode displacement due to brain shift and electrode bending. Moreover, the method may be used to improve accurate localization of MER location within the STN in post-operative recordings, which is of high importance in research studies. To the best of our knowledge, the presented model represents the first fully automatic method to fit anatomical model to recorded MER signals and as such, it may contribute to ongoing improvements in the field of stereotactic surgery.

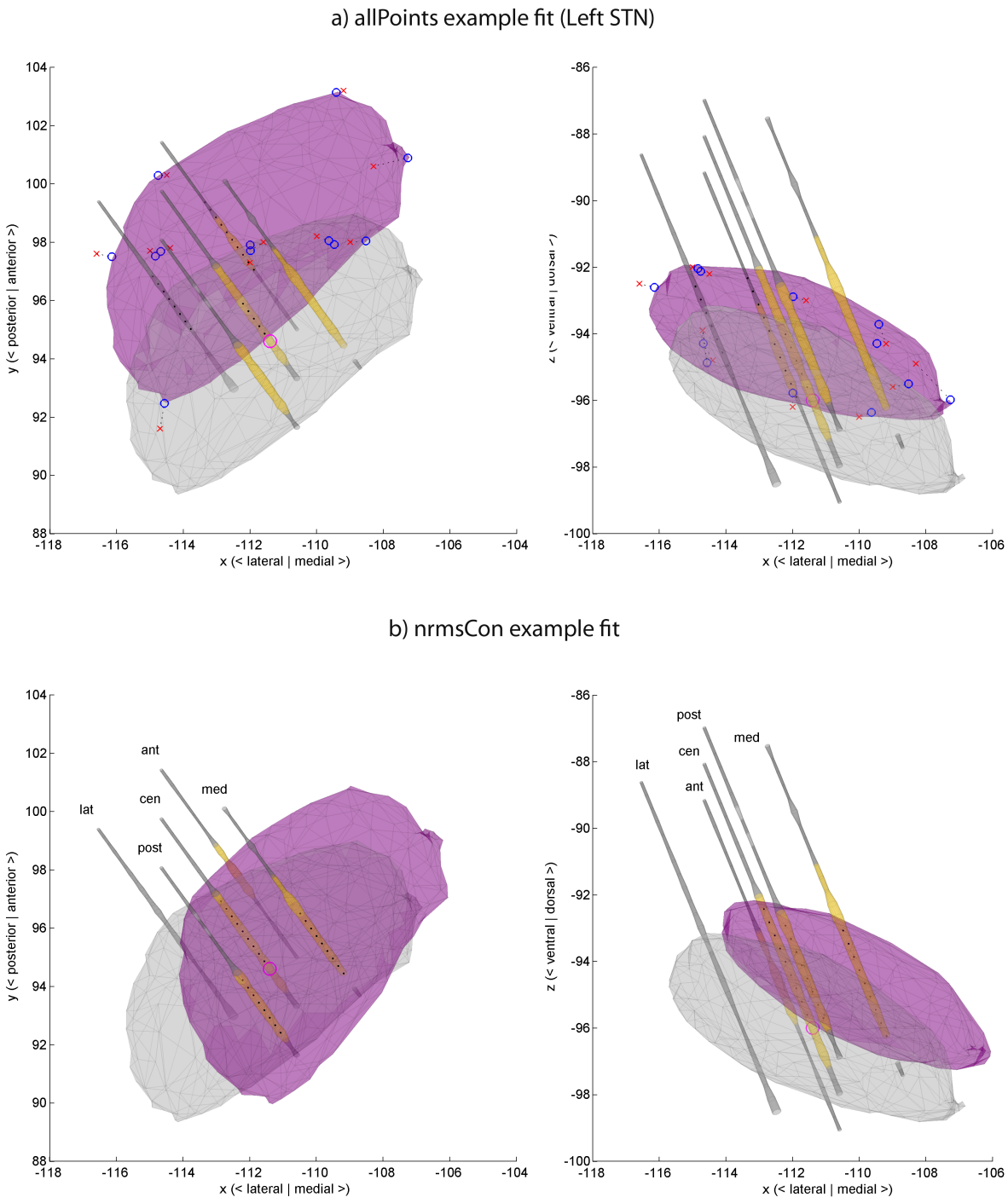

Figure 6.9: Examples of model fit using the *allPoints* method based on characteristic MRI points (top) and the proposed *nrmsCon* method based solely on electrophysiology (b - bottom) on a left (Sin) STN five-electrode trajectory. The model position after fitting is shown in purple, while the initial position (*target* method) is shown in grey. The width of the five microelectrode trajectories denotes the NRMS value, while colors denote manual labels: STN in yellow, other nuclei in grey. MER positions inside the resulting model are denoted by black points, target position shown as the red circle.

Chapter 7

Summary and perspective

In the thesis, I suggested and validated several machine learning models applied to micro and macro-electrode signals, obtained during the DBS surgery. The methods are aimed at the improvement of data preprocessing and analysis procedures both in the clinical and research setting. More accurate evaluation and elimination of errors, inherent to any real-world data, may contribute to deepening of our understanding of the disease principles, as well as to bring direct benefit to the patients in the form of increased efficacy of the treatment methods.

7.1 Thesis achievements

Scientific contribution of this thesis is represented by the following achievements:

- A method for detection of PD tremor onset from neuronal activity recorded through the stimulation electrodes has been developed and presented in Chapter 3. The algorithm achieved good classification accuracy on data from a group of PD patients and the results suggest, that an adaptive DBS device, modifying stimulation parameters to current state of the patient is feasible. The research was published in the Journal of Neuroscience Methods, IF(2012)=2.1 (Bakstein et al. 2012). Subsequent analysis, based on suggestion from the aforementioned paper was published later in Biomedical Signal Processing and Control IF(2015)=1.4 ,(Camara et al. 2015) with my minor contribution.
- Three methods for classification of MER artifact have been developed and evaluated on a newly created extensive database of labeled MER data, including more than 20 hours of annotated MER signals from four DBS centers. The proposed methods significantly outperformed existing solutions (validated previously on synthetic data only) and are provided to the research community for automatic data cleaning. The results are about to be submitted to an impacted journal.

- A GUI tool *sigInspect* was developed, that allows inspection and annotation of multi-channel MER signals. The tool incorporates methods that can be used to automatically identify MER artifacts and provide annotation that can be further refined manually. The *sigInspect* is the first tool available for automatic and semi-automatic MER signal labeling and artifact detection and has the potential to be used also for other signal types, owing to the high number of configuration options. The tool is made freely available to the research community at `https://github.com/ebakstein/sigInspect`.

- A probabilistic model for classification of brain nuclei in DBS trajectories with fuzzy states has been developed and presented. The model achieved comparable performance with the state of the art methods and proved to be suitable for the task. The main advantages include the theoretical possibility to classify with accuracy beyond the sampling interval of the measured data and easy extensibility. It has been shown, that the model can be extended by the addition of prior information and should be suitable for fitting of a 3D atlas directly to the MER recordings

- In the last Chapter, a three dimensional extension of the model has been developed and evaluated, which optimizes position of an anatomical model of the subthalamic nucleus based solely on the recorded MER signals. In comparison to model based on landmarks from pre-surgical MER imaging, the proposed model achieved a considerably higher performance and the results suggest that the proposed model may be used to identify and possibly suppress the undesired inaccuracies, inherently connected with the surgery. To the best of our knowledge, this is the first fully automatic method that extracts anatomical information solely from the electrophysiological recordings and may bring significant accuracy improvement in localization of MER recordings within anatomical structures.

7.2 Further work

In my further work, I would especially like to develop the concept of the probabilistic model for fitting of anatomical models. I consider the following research directions especially promising:

- As suggested in the discussion of Chapter 6, several steps may be undertaken to verify and possibly increase accuracy of the resulting transformation: i) addition of prior distribution on different transformation parameters, which has been already tested in case of the 1D model version in Chapter 5 and proved to be viable and ii) combination of the proposed MER-based fitting with additional information from the pre or post-surgical imagery. Both these aforementioned points may contribute to higher accuracy in both

the clinical and research scenarios and seem very promising and feasible based on the presented research, as well as additional pilot results.

- Although the model from Chapter 6 was presented in connection with electrophysiology, my interest is also the use of the model of other spatial data that can be represented by probability distributions. I would like to investigate the possibility to use the model for fitting of three-dimensional surface models to medical imaging techniques, such as the magnetic resonance imaging or computed tomography. I believe that models based on the smooth state transition are in principle generalizable to other data types and different tasks and I would like to investigate the possibilities.

- Another future task is to convert the procedure for electrophysiology-based anatomical model fitting from Chapter 6 into a software tool, available to the DBS community. A possible way is also integration to existing and well accepted tools, such as the Lead DBS (Horn et al. 2014).

7.3 List of candidate's publications

Impact factors refer to the year of publication. Citation counts according to ISI Web of Science (WoS) are valid as of Sept. 28th 2016. Author's participation equal unless other otherwise stated.

7.3.1 Publications related to the thesis topic

Journals with impact factor

- E. Bakstein, J. Burgess, K. Warwick, V. Ruiz, T. Aziz, and J. Stein (2012). "Parkinsonian tremor identification with multiple local field potential feature classification". In: *Journal of Neuroscience Methods* 209.2, pp. 320–330. DOI: 10.1016/j.jneumeth.2012.06.027
Author's participation: 60%, IF=2.1, 6 WoS citations.

- C. Camara, P. Isasi, K. Warwick, V. Ruiz, T. Aziz, J. Stein, and E. Bakstein (2015). "Resting tremor classification and detection in Parkinson's disease patients". In: *Biomedical Signal Processing and Control* 16, pp. 88–97. DOI: 10.1016/j.bspc.2014.09.006
Author's participation: 5%, IF=1.5, 5 WoS citations.

- M. Macas, L. Lhotska, E. Bakstein, D. Novak, J. Wild, T. Sieger, P. Vostatek, and R. Jech (2012). "Wrapper feature selection for small sample size data driven by complete error estimates". In: *Computer Methods and Programs in Biomedicine* 108.1, pp. 138–150. DOI:

10.1016/j.cmpb.2012.02.006
Author's participation: 5%, IF=1.6, WoS 8 citations.

Other ISI or Scopus indexed publications

- E. Bakstein, T. Sieger, D. Novak, and R. Jech (2016). "Probabilistic Model of Neuronal Background Activity in Deep Brain Stimulation Trajectories". In: *Information Technology in Bio- and Medical Informatics - 7th International Conference, (ITBAM)*
  Author's participation: 60%

- E. Bakstein, J. Schneider, T. Sieger, D. Novak, J. Wild, and R. Jech (2015). "Supervised segmentation of microelectrode recording artifacts using power spectral density". In: *Proc. of 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, At Milano, Italy.* Vol. 2015-Novem. IEEE, pp. 1524–1527. DOI: 10.1109/EMBC.2015.7318661
  Author's participation: 40%

- E. Bakstein, K. Warwick, J Burgess, O Stavdahl, and T Aziz (2010). "Features for detection of Parkinson's disease tremor from local field potentials of the subthalamic nucleus". In: *Proc. IEEE 9th Int Cybernetic Intelligent Systems (CIS) Conf*, pp. 1–6. DOI: 10.1109/UKRICIS.2010.5898092
  Author's participation: 80%

- D. Novák, F. Albert, E. Cirugueda-Roldán, T. Sieger, E. Bakštein, J. Wild, D. Cuesta, and R. Jech (2012). "Discrimination of Deep Brain Nuclei using Regularity Measures". In: *Proceedings of the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society - Abstract Book*, p. 493
  Author's participation: 4%

- J. Wild, D. Novak, E. Bakstein, T. Sieger, and R. Jech (2010). "Automatic Nuclei Detection During Parkinson's Stereotactic Neurosurgery". In: *Proceedings of Biosignal 2010: Analysis of Biomedical Signals and Images.* Vol. 20. Brno, Czech Republic: Brno University of Technology, pp. 455–458
  Author's participation: 5%

Other publications

- E. Bakstein (2011). "Modular Visualization System for Microelectrode Recordings Obtained During the Deep Brain Surgery". In: *POSTER 2011 - 15th International Student Conference on Electrical Engineering.* Prague: CTU, Faculty of Electrical Engineering,

pp. 1–4
Author's participation: 100%

7.3.2 Publications unrelated to thesis topic

Journals with impact factor

- F. Spaniel, E. Bakstein, J. Anyz, J. Hlinka, T. Sieger, J. Hrdlicka, N. Gornerova, and C. Hoschl (2016). "Relapse in schizophrenia: definitively not a bolt from the blue". In: *Neuroscience Letters* S0304-3940.16, pp. 30265–8. DOI: 10.1016/j.neulet.2016.04.044
Author's participation: 20%, IF(2015)=2.1

- P. Mikolas, T. Melicher, A. Skoch, M. Matejka, A. Slovakova, E. Bakstein, T. Hajek, and F. Spaniel (2016). "Connectivity of the anterior insula differentiates participants with first-episode schizophrenia spectrum disorders from controls: a machine-learning study". In: *Psychological Medicine*. cited By 0; Article in Press, pp. 1–10. DOI: 10.1017/S0033291716000878
Author's participation: 5%, IF(2015)=5.5

References

Abosch, A., D. Lanctin, I. Onaran, L. Eberly, M. Spaniol, and N. F. Ince (2012). "Long-term recordings of local field potentials from implanted deep brain stimulation electrodes". In: *Neurosurgery* 71.4, pp. 804–814. DOI: 10.1227/NEU.0b013e3182676b91.

Abosch, A., L. Timmermann, S. Bartley, H. G. Rietkerk, D. Whiting, P. J. Connolly, D. Lanctin, and M. I. Hariz (2013). "An international survey of deep brain stimulation procedural steps". In: *Stereotactic and Functional Neurosurgery* 91.1, pp. 1–11. DOI: 10.1159/000343207.

Aboy, M. and J. H. Falkenberg (2006). "An automatic algorithm for stationary segmentation of extracellular microelectrode recordings". In: *Medical and Biological Engineering and Computing* 44.6, pp. 511–515. DOI: 10.1007/s11517-006-0052-2.

Alesch, F., M. M. Pinter, R. J. Helscher, L. Fertl, A. L. Benabid, and W. T. Koos (1995). "Stimulation of the ventral intermediate thalamic nucleus in tremor dominated Parkinson's disease and essential tremor". In: *Acta Neurochirurgica* 136 (1). 10.1007/BF01411439, pp. 75–81.

Amtage, F., K. Henschel, B. Schelter, J. Vesper, J. Timmer, C. H. Lucking, and B. Hellwig (2008). "Tremor-correlated neuronal activity in the subthalamic nucleus of Parkinsonian patients". In: *Neuroscience Letters* 442.3, pp. 195–9. DOI: S0304-3940(08)00910-5[pii] 10.1016/j.neulet.2008.06.087.

Anderson, J. C. (1982). "Improved zerocrossing method enhances digital speech". In: *Everything Designers Need* 2.7, pp. 171–174.

Bakay, R. A. E. and A. P. Smith (2011). "Deep Brain Stimulation : Complications and Attempts at Avoiding Them". In: *The Open Neurosurgery Journal* 4.Suppl 1-M4, pp. 42–52. DOI: 10.2174/1876529701104010042.

Baker, M. G. and L. Graham (2004). "The journey: Parkinson's disease". In: *BMJ* 329.7466, pp. 611–4. DOI: 10.1136/bmj.329.7466.611329/7466/611[pii].

Bakstein, E. (2011). "Modular Visualization System for Microelectrode Recordings Obtained During the Deep Brain Surgery". In: *POSTER 2011 - 15th International Student Conference on Electrical Engineering*. Prague: CTU, Faculty of Electrical Engineering, pp. 1–4.

Bakstein, E., K. Warwick, J Burgess, O Stavdahl, and T Aziz (2010). "Features for detection of Parkinson's disease tremor from local field potentials of the subthalamic nucleus". In: *Proc.*

IEEE 9th Int Cybernetic Intelligent Systems (CIS) Conf, pp. 1–6. DOI: 10.1109/UKRICIS.2010.5898092.

Bakstein, E., J. Burgess, K. Warwick, V. Ruiz, T. Aziz, and J. Stein (2012). "Parkinsonian tremor identification with multiple local field potential feature classification". In: *Journal of Neuroscience Methods* 209.2, pp. 320–330. DOI: 10.1016/j.jneumeth.2012.06.027.

Bakstein, E., J. Schneider, T. Sieger, D. Novak, J. Wild, and R. Jech (2015). "Supervised segmentation of microelectrode recording artifacts using power spectral density". In: *Proc. of 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, At Milano, Italy.* Vol. 2015-Novem. IEEE, pp. 1524–1527. DOI: 10.1109/EMBC.2015.7318661.

Bakstein, E., T. Sieger, D. Novak, and R. Jech (2016). "Probabilistic Model of Neuronal Background Activity in Deep Brain Stimulation Trajectories". In: *Information Technology in Bio- and Medical Informatics - 7th International Conference, (ITBAM).*

Benabid, A. L., P. Pollak, D. Gao, D. Hoffmann, P. Limousin, E. Gay, I. Payen, and A. Benazzouz (1996). "Chronic electrical stimulation of the ventralis intermedius nucleus of the thalamus as a treatment of movement disorders." eng. In: *J Neurosurg* 84.2, pp. 203–214. DOI: 10.3171/jns.1996.84.2.0203.

Benabid, A. L. (2003). "Deep brain stimulation for Parkinson's disease". In: *Current Opinion in Neurobiology* 13.6, pp. 696–706. DOI: 10.1016/j.conb.2003.11.001.

Bériault, S., Y. Xiao, L. Bailey, D. L. Collins, A. F. Sadikot, and G. B. Pike (2012). "Towards computer-assisted deep brain stimulation targeting with multiple active contacts." In: *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention.* Vol. 15. Pt 1, pp. 487–94.

Brazier, M. and J. Casby (1952). "Crosscorrelation and autocorrelation studies of electroencephalographic potentials". In: *Electroencephalography and Clinical Neurophysiology* 4, pp. 201–211.

Brown, P. (2003). "Oscillatory nature of human basal ganglia activity: relationship to the pathophysiology of Parkinson's disease". In: *Mov Disord* 18.4, pp. 357–63. DOI: 10.1002/mds.10358.

Burgess, J. G., K. Warwick, V. Ruiz, M. N. Gasson, T. Z. Aziz, J. S. Brittain, and J. Stein (2010). "Identifying tremor-related characteristics of basal ganglia nuclei during movement in the Parkinsonian patient". eng. In: *Parkinsonism Relat Disord* 16.10, pp. 671–5. DOI: 10.1016/j.parkreldis.2010.08.025.

Butson, C. R., S. E. Cooper, J. M. Henderson, and C. C. McIntyre (2007). "Patient-specific analysis of the volume of tissue activated during deep brain stimulation". In: *NeuroImage* 34.2, pp. 661–670. DOI: 10.1016/j.neuroimage.2006.09.034. arXiv: NIHMS150003.

Cagnan, H., K. Dolan, X. He, M. F. Contarino, R. Schuurman, P. van den Munckhof, W. J. Wadman, L. Bour, and H. C. F. Martens (2011). "Automatic subthalamic nucleus detection from microelectrode recordings based on noise level and neuronal activity." eng. In: *J Neural Eng* 8.4, p. 046006. DOI: 10.1088/1741-2560/8/4/046006.

Camara, C., P. Isasi, K. Warwick, V. Ruiz, T. Aziz, J. Stein, and E. Bakstein (2015). "Resting tremor classification and detection in Parkinson's disease patients". In: *Biomedical Signal Processing and Control* 16, pp. 88–97. DOI: 10.1016/j.bspc.2014.09.006.

Carr, J (2002). "Tremor in Parkinson's disease". In: *Parkinsonism Relat Disord* 8.4, pp. 223–234.

Castro, F., C. Pollo, R. Meuli, P. Maeder, O. Cuisenaire, M. Cuadra, J.-G. Villemure, and J.-P. Thiran (2006). "A Cross Validation Study of Deep Brain Stimulation Targeting: From Experts to Atlas-Based, Segmentation-Based and Automatic Registration Algorithms". In: *IEEE Transactions on Medical Imaging* 25.11, pp. 1440–1450. DOI: 10.1109/TMI.2006.882129.

Chakravarty, M. M., G. Bertrand, C. P. Hodge, A. F. Sadikot, and D. L. Collins (2006). "The creation of a brain atlas for image guided neurosurgery using serial histological data". In: *NeuroImage* 30.2, pp. 359–376. DOI: 10.1016/j.neuroimage.2005.09.041.

Ciaccio, E. J., S. M. Dunn, and M. Akay (1994). "Biosignal Pattern-Recognition and Interpretation Systems". In: *Ieee Engineering in Medicine and Biology Magazine* 13.2, pp. 269–.

Daniluk, S., K. G Davies, S. A. Ellias, P. Novak, and J. M. Nazzaro (2010). "Assessment of the variability in the anatomical position and size of the subthalamic nucleus among patients with advanced Parkinson's disease using magnetic resonance imaging." eng. In: *Acta Neurochir (Wien)* 152.2, 201–10; discussion 210. DOI: 10.1007/s00701-009-0514-z.

Denker, M., S. Roux, M. Timme, A. Riehle, and S. Gr n (2007). "Phase synchronization between LFP and spiking activity in motor cortex during movement preparation". In: *Neurocomputing* 70.10-12, pp. 2096–2101.

Derya Ubeyli, E. (2009). "Statistics over features: EEG signals analysis". In: *Computers in Biology and Medicine* 39.8, pp. 733–41. DOI: S0010-4825(09)00107-3[pii]10.1016/j.compbiomed.2009.06.001.

Deuschl, G., P. Bain, and M. Brin (1998). "Consensus statement of the Movement Disorder Society on Tremor. Ad Hoc Scientific Committee". In: *Mov Disord* 13.3, pp. 2–23.

D'Haese, P. F., S. Pallavaram, R. Li, M. S. Remple, C. Kao, J. S. Neimat, P. E. Konrad, and B. M. Dawant (2012). "CranialVault and its CRAVE tools: A clinical computer assistance system for deep brain stimulation (DBS) therapy". In: *Medical Image Analysis* 16.3, pp. 744–753. DOI: 10.1016/j.media.2010.07.009. arXiv: NIHMS150003.

Dolan, K., H. C. F. Martens, P. R. Schuurman, and L. J. Bour (2009). "Automatic noise-level detection for extra-cellular micro-electrode recordings". In: *Medical and Biological Engineering and Computing* 47.7, pp. 791–800. DOI: 10.1007/s11517-009-0494-4.

Duchin, Y., A. Abosch, E. Yacoub, G. Sapiro, and N. Harel (2012). "Feasibility of using ultra-high field (7 T) MRI for clinical surgical targeting". In: *PLoS ONE* 7.5. DOI: 10.1371/journal.pone.0037328.

Egert, U., T. Knott, C. Schwarz, M. Nawrot, a. Brandt, S. Rotter, and M. Diesmann (2002). "MEA-Tools: An open source toolbox for the analysis of multi-electrode data with MATLAB". In: *Journal of Neuroscience Methods* 117.1, pp. 33–42. DOI: 10.1016/S0165-0270(02)00045-6.

Essert, C., C. Haegelen, F. Lalys, A. Abadie, and P. Jannin (2012). "Automatic computation of electrode trajectories for Deep Brain Stimulation: A hybrid symbolic and numerical approach". In: *International Journal of Computer Assisted Radiology and Surgery* 7.4, pp. 517–532. DOI: 10.1007/s11548-011-0651-8.

European Parkinson's Disease Association (2016). online, http://www.epda.eu.com/.

Fahn, S (2006). "Levodopa in the treatment of Parkinson s disease". In: *Journal of Neural Transmission* 71, pp. 1–15.

Falkenberg, J. H. and J. McNames (2003). "Segmentation of extracellular microelectrode recordings with equal power". In: *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* Vol. 3, pp. 2475–2478.

Falkenberg, J. H., J. McNames, and K. J. Burchiel (2006). "Automatic microelectrode recording analysis and visualization of the globus pallidus interna and stereotactic trajectory." eng. In: *Stereotact Funct Neurosurg* 84.1, pp. 28–34. DOI: 10.1159/000093720.

Farris, S. and M. Giroux (2011). "Deep brain stimulation: a review of the procedure and the complications." eng. In: *JAAPA* 24.2, pp. 39–40, 42–5.

Forstmann, B. U., M. C. Keuken, S. Jahfari, P. L. Bazin, J. Neumann, A. Sch??fer, A. Anwander, and R. Turner (2012). "Cortico-subthalamic white matter tract strength predicts interindividual efficacy in stopping a motor response". In: *NeuroImage* 60.1, pp. 370–375. DOI: 10.1016/j.neuroimage.2011.12.044.

Gligorijevic, I., M. Welkenhuysen, D. Prodanov, S. Musa, B. Nuttin, W. Eberle, C. Bartic, and S. V. Huffel (2009). "Neural signal analysis and artifact removal in single and multichannel in-vivo deep brain recordings". In: *Neurosurgery*, pp. 8–11.

Goetz, C. G. et al. (2007). "Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): Process, format, and clinimetric testing plan". In: *Movement Disorders* 22.1, pp. 41–47. DOI: 10.1002/mds.21198.

Gorgulho, A., A. A. F. De Salles, L. Frighetto, and E. Behnke (2005). "Incidence of hemorrhage associated with electrophysiological studies performed using macroelectrodes and

microelectrodes in functional neurosurgery." eng. In: *J Neurosurg* 102.5, pp. 888–896. DOI: 10.3171/jns.2005.102.5.0888.

Gross, R. E., P. Krack, M. C. Rodriguez-Oroz, A. R. Rezai, and A.-L. Benabid (2006). "Electrophysiological mapping for the implantation of deep brain stimulators for Parkinson's disease and tremor." eng. In: *Mov Disord* 21 Suppl 14, S259–S283. DOI: 10.1002/mds.20960.

Guarnizo, C., A. a. Orozco, and G. Castcllanos (2008). "Microelectrode signals segmentation using stationary wavelet transform". In: *BioMedical Engineering and Informatics: New Development and the Future - Proceedings of the 1st International Conference on BioMedical Engineering and Informatics, BMEI 2008* 2, pp. 450–454. DOI: 10.1109/BMEI.2008.363.

Guillen, P, F. Martinez-de Pison, R. Sanchez, M. Argaez, and L. Velazquez (2011). "Characterization of subcortical structures during deep brain stimulation utilizing support vector machines". In: *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Vol. 6. IEEE, pp. 7949–7952. DOI: 10.1109/IEMBS.2011.6091960.

Guo, T., K. W. Finnis, A. G. Parrent, and T. M. Peters (2005). "Development and application of functional databases for planning deep-brain neurosurgical procedures". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 3749 LNCS, pp. 835–842.

Halpern, C. H., S. F. Danish, G. H. Baltuch, and J. L. Jaggi (2008). "Brain shift during deep brain stimulation surgery for Parkinson's disease." eng. In: *Stereotact Funct Neurosurg* 86.1, pp. 37–43. DOI: 10.1159/000108587.

Hammerla, N. Y. and T. Plötz (2015). "Let's (not) Stick Together: Pairwise Similarity Biases Cross-Validation in Activity Recognition". In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 1041–1051. DOI: 10.1145/2750858.2807551.

Hammond, C., H. Bergman, and P. Brown (2007). "Pathological synchronization in Parkinson's disease: networks, models and treatments". In: *Trends Neurosci* 30.7, pp. 357–64. DOI: S0166-2236(07)00123-3[pii]10.1016/j.tins.2007.05.004.

Hariz, M. I. and H. Fodstad (1999). "Do microelectrode techniques increase accuracy or decrease risks in pallidotomy and deep brain stimulation? A critical review of the literature." eng. In: *Stereotact Funct Neurosurg* 72.2-4, pp. 157–169.

Harris, K. D., D. A. Henze, J. Csicsvari, H. Hirase, and G Buzsáki (2000). "Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements." In: *Journal of neurophysiology* 84.1, pp. 401–14.

Horn, A. and A. a. Kühn (2014). "Lead-DBS: A toolbox for deep brain stimulation electrode localizations and visualizations". In: *NeuroImage* 107, pp. 127–135. DOI: 10.1016/j.neuroimage.2014.12.002.

Horn Andereas, e. a. *The Lead-DBS project web page.*

Hornik, K., M. Stinchcombe, and H. White (1989). "Multilayer Feedforward Networks Are Universal Approximators". In: *Neural Networks* 2.5, pp. 359–366.

Human Brain Stimulation and Electrophysiology Lab, Department of Psychology, University of Texas at Austin (2012). online, http://brainstim.psy.utexas.edu.

Hutchison, W. D., A. M. Lozano, R. R. Tasker, A. E. Lang, and J. O. Dostrovsky (1997). "Identification and characterization of neurons with tremor-frequency activity in human globus pallidus". In: *Exp Brain Res* 113.3, pp. 557–63.

Ivan, M. E., J. Yarlagadda, A. P. Saxena, A. J. Martin, P. A. Starr, K. W. Sootsman, and P. S. Larson (2014). "Brain shift during bur hole-based procedures using interventional MRI". In: *Neurosurgery* 121, pp. 149–160. DOI: 10.3171/2014.3.JNS121312.

Jackson, J., C. K. Young, B. Hu, and B. H. Bland (2008). "High frequency stimulation of the posterior hypothalamic nucleus restores movement and reinstates hippocampal-striatal theta coherence following haloperidol-induced catalepsy". In: *Experimental Neurology* 213.1, pp. 210–9. DOI: S0014-4886(08)00252-5[pii]10.1016/j.expneurol.2008.06.007.

Jankovic, J (2008). "Parkinson's disease: clinical features and diagnosis". In: *Journal of Neurology, Neurosurgery & Psychiatry* 79.4, pp. 368–376. DOI: 10.1136/jnnp.2007.131045.

Keuken, M. C. et al. (2014). "Quantifying inter-individual anatomical variability in the subcortex using 7T structural MRI". In: *NeuroImage* 94, pp. 40–46. DOI: 10.1016/j.neuroimage.2014.03.032.

Krack, P. et al. (2003). "Five-year follow-up of bilateral stimulation of the subthalamic nucleus in advanced Parkinson's disease." eng. In: *N Engl J Med* 349.20, pp. 1925–1934. DOI: 10.1056/NEJMoa035275.

Krauth, A., R. Blanc, A. Poveda, D. Jeanmonod, A. Morel, and G. Szekely (2010). "A mean three-dimensional atlas of the human thalamus: Generation from multiple histological data". In: *NeuroImage* 49.3, pp. 2053–2062. DOI: 10.1016/j.neuroimage.2009.10.042.

Kuhn, A. A., T. Trottenberg, A. Kivi, A. Kupsch, G. H. Schneider, and P. Brown (2005). "The relationship between local field potential and neuronal discharge in the subthalamic nucleus of patients with Parkinson's disease". In: *Experimental Neurology* 194, pp. 212–220.

Kuhn, A. A. et al. (2008). "High-frequency stimulation of the subthalamic nucleus suppresses oscillatory beta activity in patients with Parkinson's disease in parallel with improvement in motor performance". In: *J Neurosci* 28.24, pp. 6165–73. DOI: 10.1523/JNEUROSCI.0282-08.2008.

Lalys, F., C. Haegelen, M. Mehri, S. Drapier, M. Verin, and P. Jannin (2013). "Anatomo-clinical atlases correlate clinical data and electrode contact coordinates: Application to subthalamic deep brain stimulation". In: *Journal of Neuroscience Methods* 212.2, pp. 297–307. DOI: 10.1016/j.jneumeth.2012.11.002.

Lalys, F., C. Haegelen, T. D'Albis, and P. Jannin (2014). "Analysis of electrode deformations in deep brain stimulation surgery". In: *International Journal of Computer Assisted Radiology and Surgery* 9.1, pp. 107–117. DOI: 10.1007/s11548-013-0911-x.

Lawhern, V., W. D. Hairston, and K. Robbins (2013). "DETECT: a MATLAB toolbox for event detection and identification in time series, with applications to artifact detection in EEG signals." In: *PloS one* 8.4, e62944. DOI: 10.1371/journal.pone.0062944.

Lemstra, A. W., L. Verhagen Metman, J. I. Lee, P. M. Dougherty, and F. A. Lenz (1999). "Tremor-frequency (3-6 Hz) activity in the sensorimotor arm representation of the internal segment of the globus pallidus in patients with Parkinson's disease". In: *Neuroscience Letters* 267.2, pp. 129–32. DOI: S0304-3940(99)00343-2[pii].

Levy, R., W. D. Hutchison, A. M. Lozano, and J. O. Dostrovsky (2002). "Synchronized neuronal discharge in the basal ganglia of parkinsonian patients is limited to oscillatory activity." eng. In: *J Neurosci* 22.7, pp. 2855–2861. DOI: 20026193.

Limousin, P., P. Krack, P. Pollak, A. Benazzouz, C. Ardouin, D. Hoffmann, and A. L. Benabid (1998). "Electrical stimulation of the subthalamic nucleus in advanced Parkinson's disease." eng. In: *N Engl J Med* 339.16, pp. 1105–1111. DOI: 10.1056/NEJM199810153391603.

Liu, X., H. L. Ford-Dunn, G. N. Hayward, D. Nandi, R. C. Miall, T. Z. Aziz, and J. F. Stein (2002). "The oscillatory activity in the Parkinsonian subthalamic nucleus investigated using the macro-electrodes for deep brain stimulation". In: *Clinical Neurophysiology* 113.11, pp. 1667–72.

Loukas, C. and P. Brown (2004). "Online prediction of self-paced hand-movements from subthalamic activity using neural networks in Parkinson s disease". In: *Journal of Neuroscience Methods* 137.2, pp. 193–205.

Lourens, M., H. Meijer, M. Contarino, P. van den Munckhof, P. Schuurman, S. van Gils, and L. Bour (2013). "Functional neuronal activity and connectivity within the subthalamic nucleus in Parkinson's disease". In: *Clinical Neurophysiology* 124.5, pp. 967–981. DOI: 10.1016/j.clinph.2012.10.018.

Lujan, J. L., A. M. Noecker, C. R. Butson, S. E. Cooper, B. L. Walter, J. L. Vitek, and C. C. McIntyre (2009). "Automated 3-Dimensional Brain Atlas Fitting to Microelectrode Recordings from Deep Brain Stimulation Surgeries". In: *Stereotactic and Functional Neurosurgery* 87.4, pp. 229–240. DOI: 10.1159/000225976.

Macas, M., L. Lhotska, E. Bakstein, D. Novak, J. Wild, T. Sieger, P. Vostatek, and R. Jech (2012). "Wrapper feature selection for small sample size data driven by complete error estimates". In: *Computer Methods and Programs in Biomedicine* 108.1, pp. 138–150. DOI: 10.1016/j.cmpb.2012.02.006.

Magarinos-Ascone, C. M., R. Figueiras-Mendez, C. Riva-Meana, and A. Cordoba-Fernandez (2000). "Subthalamic neuron activity related to tremor and movement in Parkinson's disease". In: *Eur J Neurosci* 12.7, pp. 2597–607.

Marceglia, S., M. Fiorio, G. Foffani, S. Mrakic-Sposta, M. Tiriticco, M. Locatelli, E. Caputo, M. Tinazzi, and A. Priori (2009). "Modulation of beta oscillations in the subthalamic area during action observation in Parkinson's disease". In: *Neuroscience* 161.4, pp. 1027–36. DOI: S0306-4522(09)00612-5[pii]10.1016/j.neuroscience.2009.04.018.

Marsden, C. D. (1994). "Problems with long-term levodopa therapy for Parkinson's disease." eng. In: *Clinical neuropharmacology* 17 Suppl 2, S32–44.

Marsden, J. F., P. Ashby, P. Limousin-Dowsey, J. C. Rothwell, and P. Brown (2000). "Coherence between cerebellar thalamus, cortex and muscle in man - Cerebellar thalamus interactions". In: *Brain* 123, pp. 1459–1470.

Marsden, J. F., P. Limousin-Dowsey, P. Ashby, P. Pollak, and P. Brown (2001). "Subthalamic nucleus, sensorimotor cortex and muscle interrelationships in Parkinson's disease". In: *Brain* 124, pp. 378–388.

Martin, A. J., P. S. Larson, J. L. Ostrem, W. K. Sootsman, P. Talke, O. M. Weber, N. Levesque, J. Myers, and P. A. Starr (2005). "Placement of deep brain stimulator electrodes using real-time high-field interventional magnetic resonance imaging". In: *Magnetic Resonance in Medicine* 54.5, pp. 1107–1114. DOI: 10.1002/mrm.20675.

Martinez, J., C. Pedreira, M. J. Ison, and R. Quian Quiroga (2009). "Realistic simulation of extracellular recordings". In: *Journal of Neuroscience Methods* 184, pp. 285–293. DOI: 10.1016/j.jneumeth.2009.08.017.

Masuda, T., H. Miyano, and T. Sadoyama (1982). "The measurement of muscle fiber conduction velocity using a gradient threshold zero-crossing method". In: *IEEE Transactions on Biomedical Engineering* 29.10, pp. 673–8. DOI: 10.1109/TBME.1982.324859.

McIntyre, C. C., A. Chaturvedi, R. R. Shamir, and S. F. Lempka (2015). "Engineering the next generation of clinical deep brain stimulation technology". In: *Brain Stimulation* 8.1, pp. 21–26. DOI: 10.1016/j.brs.2014.07.039. arXiv: 15334406.

Medtronics online catalog (2012). online, http://professional.medtronic.com/pt/neuro/dbs-md/prod/index.htm.

Mesa, L. a. Camuñas and R. Q. Quiroga (2013). "A detailed and fast model of extracellular recordings." In: *Neural computation* 25, pp. 1191–212. DOI: 10.1162/NECO_a_00433.

Michmizos, K. P., D. Sakas, and K. S. Nikita (2012). "Parameter identification for a local field potential driven model of the Parkinsonian subthalamic nucleus spike activity". In: *Neural Networks* 36, pp. 146–156. DOI: 10.1016/j.neunet.2012.10.003.

Mikolas, P., T. Melicher, A. Skoch, M. Matejka, A. Slovakova, E. Bakstein, T. Hajek, and F. Spaniel (2016). "Connectivity of the anterior insula differentiates participants with first-

episode schizophrenia spectrum disorders from controls: a machine-learning study". In: *Psychological Medicine*. cited By 0; Article in Press, pp. 1–10. DOI: 10.1017/S0033291716000878.

Miocinovic, S, A. M. Noecker, C. B. Maks, C. R. Butson, and C. C. McIntyre (2007). "Cicerone: stereotactic neurophysiological recording and deep brain stimulation electrode placement software system". In: *Operative Neuromodulation: Volume 2: Neural Networks Surgery*. Ed. by D. E. Sakas and B. A. Simpson. Vienna: Springer Vienna, pp. 561–567. DOI: 10.1007/978-3-211-33081-4_65.

Moran, A., H. Bergman, Z. Israel, and I. Bar-Gad (2008). "Subthalamic nucleus functional organization revealed by parkinsonian neuronal oscillations and synchrony". In: *Brain* 131.Pt 12, pp. 3395–409. DOI: awn270[pii]10.1093/brain/awn270.

Moran, A., I. Bar-Gad, H. Bergman, and Z. Israel (2006). "Real-time refinement of subthalamic nucleus targeting using Bayesian decision-making on the root mean square measure." In: *Mov Disord* 21.9, pp. 1425–1431. DOI: 10.1002/mds.20995.

Morel, A. (2007). *Stereotactic atlas of the human thalamus and basal ganglia*. CRC Press,Taylor & Francis Group.

Morrison, S., G. Kerr, K. M. Newell, and P. A. Silburn (2008). "Differential time- and frequency-dependent structure of postural sway and finger tremor in Parkinson's disease". In: *Neurosci Lett* 443.3, pp. 123–8. DOI: S0304-3940(08)01068-9[pii]10.1016/j.neulet.2008.07.071.

Nimsky, C., O. Ganslandt, S. Cerny, P. Hastreiter, G. Greiner, and R. Fahlbusch (2000). "Quantification of, visualization of, and compensation for brain shift using intraoperative magnetic resonance imaging." eng. In: *Neurosurgery* 47.5, 1070–9; discussion 1079–80.

Novak, P., S. Daniluk, S. a. Ellias, and J. M. Nazzaro (2007). "Detection of the subthalamic nucleus in microelectrographic recordings in Parkinson disease using the high-frequency (>500 hz) neuronal background." In: *Journal of neurosurgery* 106.1, pp. 175–179. DOI: 10.3171/jns.2007.106.1.175.

Novak, P., A. W. Przybyszewski, A. Barborica, P. Ravin, L. Margolin, and J. G. Pilitsis (2011). "Localization of the subthalamic nucleus in Parkinson disease using multiunit activity". In: *Journal of the Neurological Sciences* 310.1-2, pp. 44–49. DOI: 10.1016/j.jns.2011.07.027.

Nowinski, W. L., D. Belov, P. Pollak, and A. L. Benabid (2004). "A probabilistic functional atlas of the human subthalamic nucleus." In: *Neuroinformatics* 2.4, pp. 381–98. DOI: 10.1385/NI:2:4:381.

Obeso, J. A. and J. L. Lanciego (2011). "Past, present, and future of the pathophysiological model of the Basal Ganglia." In: *Frontiers in neuroanatomy* 5.July, p. 39. DOI: 10.3389/fnana.2011.00039.

Obien, M. E. J., K. Deligkaris, T. Bullmann, D. J. Bakkum, and U. Frey (2015). "Revealing neuronal function through microelectrode array recordings". In: *Frontiers in Neuroscience* 8.January, pp. 1–30. DOI: 10.3389/fnins.2014.00423.

Olanow, C. W. and W. G. Tatton (1999). "Etiology and Pathogenesis of Parkinson'S Disease". In: *Annual Review of Neuroscience* 22.1, pp. 123–144. DOI: 10.1146/annurev.neuro.22.1.123.

Pahwa, R. et al. (2006). "Practice Parameter: treatment of Parkinson disease with motor fluctuations and dyskinesia (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology." eng. In: *Neurology* 66.7, pp. 983–995.

Pallavaram, S., B. M. Dawant, M. S. Remple, J. S. Neimat, C. Kao, P. E. Konrad, and P. F. D'Haese (2010). "Effect of brain shift on the creation of functional atlases for deep brain stimulation surgery". In: *International Journal of Computer Assisted Radiology and Surgery* 5.3, pp. 221–228. DOI: 10.1007/s11548-009-0391-1.

Pan, S., S. Iplikci, K. Warwick, and T. Z. Aziz (2012). "Parkinson s Disease tremor classification A comparison between Support Vector Machines and neural networks". In: *Expert Systems with Applications* 39.12, pp. 10764–10771. DOI: 10.1016/j.eswa.2012.02.189.

Paralikar, K. J., C. R. Rao, and R. S. Clement (2009). "New approaches to eliminating common-noise artifacts in recordings from intracortical microelectrode arrays: Inter-electrode correlation and virtual referencing". In: *Journal of Neuroscience Methods* 181.1, pp. 27–35. DOI: 10.1016/j.jneumeth.2009.04.014.

Pesenti, A. et al. (2004). "The subthalamic nucleus in Parkinson's disease: power spectral density analysis of neural intraoperative signals." eng. In: *Neurol Sci* 24.6, pp. 367–374. DOI: 10.1007/s10072-003-0191-2.

Pollo, C., A. Kaelin-Lang, M. F. Oertel, L. Stieglitz, E. Taub, P. Fuhr, A. M. Lozano, A. Raabe, and M. Sch??pbach (2014). "Directional deep brain stimulation: An intraoperative double-blind pilot study". In: *Brain* 137.7, pp. 2015–2026. DOI: 10.1093/brain/awu102.

Prodoehl, J., H. Yu, D. M. Little, I. Abraham, and D. E. Vaillancourt (2008). "Region of interest template for the human basal ganglia: Comparing EPI and standardized space approaches". In: *NeuroImage* 39.3, pp. 956–965. DOI: 10.1016/j.neuroimage.2007.09.027.

Pulliam, C. L., D. a. Heldman, T. H. Orcutt, T. O. Mera, J. P. Giuffrida, and J. L. Vitek (2015). "Motion Sensor Strategies for Automated Optimization of Deep Brain Stimulation in Parkinson's disease". In: *Parkinsonism & Related Disorders* 21.4, pp. 378–382. DOI: 10.1016/j.parkreldis.2015.01.018.

Qiu, A, T Brown, B Fischl, J Ma, and M. I. Miller (2010). "Atlas Generation for Subcortical and Ventricular Structures With Its Applications in Shape Analysis". In: *IEEE Transactions on Image Processing* 19.6, pp. 1539–1547. DOI: 10.1109/TIP.2010.2042099.

Quiroga, R. Q., Z Nadasdy, and Y Ben-Shaul (2004). "Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering." In: *Neural computation* 16, pp. 1661–1687. DOI: 10.1162/089976604774201631.

Rajput, A. H. and S. Birdi (1997). "Epidemiology of Parkinson's disease". In: *Parkinsonism Relat Disord* 3.4, pp. 175–86. DOI: S1353-8020(97)00029-1[pii].

Ray, N. J. et al. (2009a). "Abnormal thalamocortical dynamics may be altered by deep brain stimulation: using magnetoencephalography to study phantom limb pain". In: *J Clin Neurosci* 16.1, pp. 32–6. DOI: 10.1016/j.jocn.2008.03.004.

Ray, N. J. et al. (2009b). "The role of the subthalamic nucleus in response inhibition: evidence from deep brain stimulation for Parkinson's disease". In: *Neuropsychologia* 47.13, pp. 2828–34. DOI: 10.1016/j.neuropsychologia.2009.06.011.

Reggia, J. A., Y. Peng, and S. Tuhrim (1993). "A Connectionist Approach to Diagnostic Problem-Solving Using Causal Networks". In: *Information Sciences* 70.1-2, pp. 27–48.

Rezai, A. R., B. H. Kopell, R. E. Gross, J. L. Vitek, A. D. Sharan, P. Limousin, and A.-L. Benabid (2006). "Deep brain stimulation for Parkinson's disease: surgical issues." eng. In: *Mov Disord* 21 Suppl 14, S197–S218. DOI: 10.1002/mds.20956.

Rivlin-Etzion, M., O. Marmor, G. Heimer, A. Raz, A. Nini, and H. Bergman (2006). "Basal ganglia oscillations and pathophysiology of movement disorders". In: *Current Opinion in Neurobiology* 16.6, pp. 629–37. DOI: S0959-4388(06)00140-1[pii]10.1016/j.conb.2006.10.002.

Rodriguez, M. C., O. J. Guridi, L. Alvarez, K. Mewes, R. Macias, J. Vitek, M. R. DeLong, and J. A. Obeso (1998). "The subthalamic nucleus and tremor in Parkinson's disease". In: *Mov Disord* 13 Suppl 3, pp. 111–8.

Rutishauser, U., E. M. Schuman, and A. N. Mamelak (2006). "Online detection and sorting of extracellularly recorded action potentials in human medial temporal lobe recordings, in vivo". In: *Journal of Neuroscience Methods* 154.1-2, pp. 204–224. DOI: 10.1016/j.jneumeth.2005.12.033. arXiv: 0604033 [q-bio].

Sadikot, A. F., M. M. Chakravarty, G. Bertrand, V. V. Rymar, F. Al-Subaie, and D. L. Collins (2011). "Creation of Computerized 3D MRI-Integrated Atlases of the Human Basal Ganglia and Thalamus." In: *Frontiers in systems neuroscience* 5.September, p. 71. DOI: 10.3389/fnsys.2011.00071.

Seifried, C., L. Weise, and R. Hartmann (2012). "Intraoperative microelectrode recording for the delineation of subthalamic nucleus topography in Parkinson's disease". In: *Brain stimulation* 5.3, pp. 378–387. DOI: 10.1016/j.brs.2011.06.002.

Shamir, R. R., A. Zaidel, L. Joskowicz, H. Bergman, and Z. Israel (2012). "Microelectrode Recording Duration and Spatial Density Constraints for Automatic Targeting of the Sub-

thalamic Nucleus". In: *Stereotactic and Functional Neurosurgery* 90.5, pp. 325–334. DOI: 10.1159/000338252.

Sieger, T. et al. (2015). "Distinct populations of neurons respond to emotional valence and arousal in the human subthalamic nucleus." In: *Proceedings of the National Academy of Sciences of the United States of America* 112.10, pp. 3116–21. DOI: 10.1073/pnas.1410709112.

Simon, S. L., P. Douglas, G. H. Baltuch, and J. L. Jaggi (2005). "Error analysis of MRI and leksell stereotactic frame target localization in deep brain stimulation surgery." eng. In: *Stereotact Funct Neurosurg* 83.1, pp. 1–5. DOI: 10.1159/000083861.

Skotte, J., N. Hjortskov, M. Essendrop, B. Schibye, and N. Fallentin (2005). "Short latency stretch reflex in human lumbar paraspinal muscles". In: *Journal of Neuroscience Methods* 145.1-2, pp. 145–50. DOI: S0165-0270(04)00440-6[pii]10.1016/j.jneumeth.2004.12.006.

Slavin, K and J Holsapple (2004). "Micro electrode Techniques Equipment Components and Systems". In: *Microelectrode Recording in Movement Disorder Surgery*. Ed. by Z. Israel and K. Burchiel. Thieme, pp. 14–27.

Spaniel, F., E. Bakstein, J. Anyz, J. Hlinka, T. Sieger, J. Hrdlicka, N. Gornerova, and C. Hoschl (2016). "Relapse in schizophrenia: definitively not a bolt from the blue". In: *Neuroscience Letters* S0304-3940.16, pp. 30265–8. DOI: 10.1016/j.neulet.2016.04.044.

Stacey, W. C., S. Kellis, B. Greger, C. R. Butson, P. R. Patel, T. Assaf, T. Mihaylova, and S. Glynn (2013). "Potential for unreliable interpretation of EEG recorded with microelectrodes". In: *Epilepsia* 54.8, pp. 1391–1401. DOI: 10.1111/epi.12202.

Steigerwald, F., M. Potter, J. Herzog, M. Pinsker, F. Kopper, H. Mehdorn, G. Deuschl, and J. Volkmann (2008). "Neuronal activity of the human subthalamic nucleus in the parkinsonian and nonparkinsonian state". In: *J Neurophysiol* 100.5, pp. 2515–24. DOI: 90574.2008[pii]10.1152/jn.90574.2008.

Taghva, A. (2010). "An automated navigation system for deep brain stimulator placement using hidden Markov models." eng. In: *Neurosurgery* 66.3 Suppl Operative, 108–17; discussion 117. DOI: 10.1227/01.NEU.0000365369.48392.E8.

Taghva, A. (2011). "Hidden Semi-Markov Models in the Computerized Decoding of Microelectrode Recording Data for Deep Brain Stimulator Placement". In: *World Neurosurgery* 75.5-6, 758–763.e4. DOI: 10.1016/j.wneu.2010.11.008.

Tagluk, M. E. and N. Sezgin (2011). "A new approach for estimation of obstructive sleep apnea syndrome". In: *Expert Systems with Applications* 38.5, pp. 5346–5351.

Tarsy, D., J. Vitek, P. Starr, and M. Okun (2008). *Deep Brain Stimulation in Neurological and Psychiatric Disorders*. Current Clinical Neurology. Humana Press.

Thani, N. B., A. Bala, G. B. Swann, and C. R. P. Lind (2011). "Accuracy of postoperative computed tomography and magnetic resonance image fusion for assessing deep brain stim-

ulation electrodes." eng. In: *Neurosurgery* 69.1, 207–14; discussion 214. DOI: 10.1227/NEU.0b013e318218c7ae.

Trottenberg, T., A. Kupsch, G. H. Schneider, P. Brown, and A. A. Kuhn (2007). "Frequency-dependent distribution of local field potential activity within the subthalamic nucleus in Parkinson's disease". In: *Exp Neurol* 205.1, pp. 287–91. DOI: S0014-4886(07)00047-7[pii] 10.1016/j.expneurol.2007.01.028.

Ubeyli, E. D. (2008). "Time-varying biomedical signals analysis with multiclass support vector machines employing Lyapunov exponents". In: *Digital Signal Processing* 18.4, pp. 646–656. DOI: DOI10.1016/j.dsp.2007.10.001.

Van Den Eeden, S. K. (2003). "Incidence of Parkinson's Disease: Variation by Age, Gender, and Race/Ethnicity". In: *American Journal of Epidemiology* 157.¿11, pp. 1015–1022. DOI: 10.1093/aje/kwg068.

Van Den Munckhof, P., M. F. Contarino, L. J. Bour, J. D. Speelman, R. M. A. De Bie, and P. R. Schuurman (2010). "Postoperative curving and upward displacement of deep brain stimulation electrodes caused by brain shift". In: *Neurosurgery* 67.1, pp. 49–53. DOI: 10.1227/01.NEU.0000370597.44524.6D.

Verhagen, R. et al. (2015). "Advanced target identification in STN-DBS with beta power of combined local field potentials and spiking activity". In: *Journal of Neuroscience Methods* 253, pp. 116–125. DOI: 10.1016/j.jneumeth.2015.06.006.

Videen, T. O., M. C. Campbell, S. D. Tabbal, M. Karimi, T. Hershey, and J. S. Perlmutter (2008). "Validation of a fiducial-based atlas localization method for deep brain stimulation contacts in the area of the subthalamic nucleus". In: *Journal of Neuroscience Methods* 168.2, pp. 275–281. DOI: 10.1016/j.jneumeth.2007.10.007.

Wagenaar, D., T. DeMarse, and S. Potter (2005). "MeaBench: A toolset for multi-electrode data acquisition and on-line analysis". In: *Conference Proceedings. 2nd International IEEE EMBS Conference on Neural Engineering, 2005.* Vol. 2005. IEEE, pp. 518–521. DOI: 10.1109/CNE.2005.1419673.

Wagenaar, D. a. and S. M. Potter (2002). "Real-time multi-channel stimulus artifact suppression by local curve fitting". In: *Journal of Neuroscience Methods* 120.2, pp. 113–120. DOI: 10.1016/S0165-0270(02)00149-8.

Wang, S., Y. Chen, M. Ding, J. Feng, J. F. Stein, T. Z. Aziz, and X. Liu (2007). "Revealing the dynamic causal interdependence between neural and muscular signals in Parkinsonian tremor". In: *Journal of the Franklin Institute* 344.33-4, pp. 180–195.

Wang, S. Y., T. Z. Aziz, J. F. Stein, and X. Liu (2005). "Time-frequency analysis of transient neuromuscular events: dynamic changes in activity of the subthalamic nucleus and forearm muscles related to the intermittent resting tremor". In: *J Neurosci Methods* 145.1-2, pp. 151–8. DOI: 10.1016/j.jneumeth.2004.12.009.

Weegink, K. J., P. a. Bellette, J. J. Varghese, P. a. Silburn, P. a. Meehan, and A. P. Bradley (2013). "Efficient Micro-electrode Recording Modeling using a Filtered Point Process". In: *arXiv*, pp. 1–22. arXiv: 1307.5250.

Weinberger, M., N. Mahant, W. D. Hutchison, A. M. Lozano, E. Moro, M. Hodaie, A. E. Lang, and J. O. Dostrovsky (2006). "Beta oscillatory activity in the subthalamic nucleus and its relation to dopaminergic response in Parkinson's disease". In: *Journal of Neurophysiology* 96.6, pp. 3248–56. DOI: 00697.2006[pii]10.1152/jn.00697.2006.

Wild, J., D. Novak, E. Bakstein, T. Sieger, and R. Jech (2010). "Automatic Nuclei Detection During Parkinson's Stereotactic Neurosurgery". In: *Proceedings of Biosignal 2010: Analysis of Biomedical Signals and Images*. Vol. 20. Brno, Czech Republic: Brno University of Technology, pp. 455–458.

Wild, J., Z. Prekopcsak, T. Sieger, D. Novak, and R. Jech (2012). "Performance comparison of extracellular spike sorting algorithms for single-channel recordings." eng. In: *J Neurosci Methods* 203.2, pp. 369–376. DOI: 10.1016/j.jneumeth.2011.10.013.

Wiley, R. G., H. Schwarzlander, and D. D. Weiner (1977). "Demodulation procedure for verywideband FM". In: *IEEE Transactions on Communications* 25, pp. 318–327.

Winestone, J. S., A. Zaidel, H. Bergman, and Z. Israel (2012). "The use of macroelectrodes in recording cellular spiking activity". In: *Journal of Neuroscience Methods* 206.1, pp. 34–39. DOI: 10.1016/j.jneumeth.2012.02.002.

Wong, S., G. H. Baltuch, J. L. Jaggi, and S. F. Danish (2009). "Functional localization and visualization of the subthalamic nucleus from microelectrode recordings acquired during DBS surgery with unsupervised machine learning." eng. In: *J Neural Eng* 6.2, p. 026006. DOI: 10.1088/1741-2560/6/2/026006.

Wu, D. F., K. Warwick, Z. Ma, M. N. Gasson, J. G. Burgess, S. Pan, and T. Z. Aziz (2010). "Prediction of Parkinson's Disease Tremor Onset Using a Radial Basis Function Neural Network Based on Particle Swarm Optimization". eng. In: *International Journal of Neural Systems* 20.2, pp. 109–116. DOI: Doi10.1142/S0129065710002292.

Wua, C.-H. et al. (2011). "Frequency recognition in an SSVEP-based brain computer interface using empirical mode decomposition and refined generalized zero-crossing". In: *Journal of Neuroscience Methods* 196.1, pp. 170–181.

Yelnik, J. et al. (2007). "A three-dimensional, histological and deformable atlas of the human basal ganglia. I. Atlas construction based on immunohistochemical and MRI data". In: *NeuroImage* 34.2, pp. 618–638. DOI: 10.1016/j.neuroimage.2006.09.026.

Zaidel, A., A. Spivak, B. Grieb, H. Bergman, and Z. Israel (2010). "Subthalamic span of beta oscillations predicts deep brain stimulation efficacy for patients with Parkinson's disease". In: *Brain* 133.Pt 7, pp. 2007–21. DOI: 10.1093/brain/awq144.

Zaidel, A., A. Spivak, L. Shpigelman, H. Bergman, and Z. Israel (2009). "Delimiting subterritories of the human subthalamic nucleus by means of microelectrode recordings and a Hidden Markov Model." In: *Mov Disord* 24.12, pp. 1785–1793. DOI: 10.1002/mds.22674.

Zonenshayn, M., D. Sterio, P. J. Kelly, A. R. Rezai, and A. Beric (2004). "Location of the active contact within the subthalamic nucleus (STN) in the treatment of idiopathic Parkinson's disease." eng. In: *Surg Neurol* 62.3, 216–25; discussion 225–6. DOI: 10.1016/j.surneu.2003.09.039.