Antti Tanner

Automatic seizure detection using a two-dimensional EEG feature space

Thesis submitted for examination for the degree of Master of Science in Technology.

Helsinki 6.9.2011

Thesis supervisor:

Prof. Risto Ilmoniemi

Thesis instructor:

D.Sc. (Tech.) Mika Särkelä

Aalto University
School of Science

AALTO UNIVERSITY
SCHOOL OF SCIENCE

ABSTRACT OF THE
MASTER'S THESIS

Author: Antti Tanner

Title: Automatic seizure detection using a two-dimensional EEG feature space

Date: 6.9.2011 Language: English Number of pages:10+61

Department of Biomedical Engineering and Computational Science

Professorship: Biomedical Engineering Code: Tfy-99

Supervisor: Prof. Risto Ilmoniemi

Instructor: D.Sc. (Tech.) Mika Särkelä

Epileptic seizures are neurological dysfunctions that are manifested in abnormal electrical activity of the brain. Behavioural correlates, such as convulsions, are sometimes associated with seizures. There are, however, seizures that do not have clear external manifestations. These non-convulsive seizures can be detected only by monitoring brain activity. Accumulating evidence suggests that non-convulsive seizures are particularly common in intensive care units (ICUs), even among patients with no prior seizures. Presence of seizures is a medical emergency that requires fast intervention.

Electroencephalogram (EEG) can be used to monitor brain's electrical activity. In EEG, potential differences are measured from different sites on the subject's scalp. Long-term measurements generate a lot of data and manually reviewing all of it is an exhausting task. There is a clear need for an automatic seizure detection method.

In this study, three methods are proposed for seizure detection. We compute instantaneous frequency and signal power from EEG and quantify the evolution of these features. The first method measures the length of the path that feature vectors create in the feature space. The second method compares the latest step to the average step. The last method encloses the background activity in a convex hull and classifies epochs that breach the hull.

The third method was found to have the best overall performance. It can potentially detect 11 out of 19 seizure patients in the database. The database consists of recordings from 179 ICU patients. Most of the false positive detections were caused by muscle artefact, other signal artefacts, or rudimentary detection logic.

The developed methods have good potential in detecting certain types of seizures. Before reporting final performance numbers, the algorithm must be complemented with a spike detection algorithm and a proper artefact detection algorithm.

Keywords: EEG, seizure, epilepsy, wavelet analysis, Hilbert transform, ICU

AALTO-YLIOPISTO
PERUSTIETEIDEN KORKEAKOULU

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä: Antti Tanner

Työn nimi: Epileptisten kohtauksien automaattinen tunnistaminen kaksiulotteisessa EEG-piirreavaruudessa

Päivämäärä: 6.9.2011 Kieli: Englanti Sivumäärä:10+61

Lääketieteellisen tekniikan ja laskennallisen tieteen laitos

Professuuri: Lääketieteellinen tekniikka Koodi: Tfy-99

Valvoja: Prof. Risto Ilmoniemi

Ohjaaja: TkT Mika Särkelä

Epileptinen kohtaus on neurologinen häiriötila, joka ilmenee aivojen epänormaalina sähköisenä toimintana. Joihinkin kohtauksiin liittyy ulkoisia merkkejä, kuten lihaskouristuksia. Kohtauksia, joihin ei liity selkeitä ulkoisia merkkejä, kutsutaan ei-konvulsiiviksi. Ne voidaan tunnistaa vain seuraamalla aivojen sähköistä toimintaa. Ei-konvulsiivisten kohtauksien on osoitettu olevan erityisen yleisiä tehohoitopotilailla – myös sellaisilla potilailla, joilla ei ole aiemmin ollut kohtauksia. Epileptinen kohtaus on pikaista interventiota vaativa vakava tila.

Aivosähkökäyrällä (elektroenkefalografia, EEG) voidaan tutkia aivojen sähköistä toimintaa. Datan läpikäynti käsin on aikaavievää, joten tehohoitoon sopivalle, automaattiselle ja reaaliaikaiselle analyysimenetelmälle on suuri tarve.

Tässä diplomityössä esitellään kolme menetelmää, jotka soveltuvat signaalipiirteiden evoluution seuraamiseen. Kultakin EEG-kanavalta määritetään kaksi piirrettä: hetkellinen taajuus ja signaalin teho. Ensimmäinen menetelmä mittaa piirreavaruuteen muodostuvan polun pituutta aikatasossa. Toinen menetelmä vertaa kutakin piirreavaruudessa otettua askelta edellisiin askeliin. Kolmannessa menetelmässä määritetään dynaamisesti edellisistä piirrevektoreista konveksi kuori ja tutkitaan kuoren ulkopuolelle osuvia piirrevektoreita.

Kolmas menetelmä osoittautui tutkimuksessa parhaaksi. Menetelmällä pystyttiin tunnistamaan 11 tietokannan 19:sta kohtauksista kärsineestä potilaasta. Tietokannassa on EEG-mittauksia 179 tehohoitopotilaalta. Suurin osa vääristä detektioista johtui EEG:ssä näkyvästä lihastoiminnasta, artefaktoista tai alkeellisesta tunnistuslogiikasta.

Menetelmän todellista suorituskykyä on liian aikaista arvioida. Menetelmää pitää täydentää EEG-piikit sekä artefaktat luotettavasti tunnistavilla algoritmeilla.

Avainsanat: EEG, epileptinen kohtaus, epilepsia, wavelet-analyysi, Hilbert-muunnos, tehohoito

Preface

This work was carried out in the Technology Research team of General Electric Healthcare Finland Oy. I would like to thank the whole research team and my colleagues at GE for introducing me to the world of clinical research and product development. In particular, I thank my instructor Mika Särkelä for his advice and for his contagious enthusiasm. I also thank my supervisor, professor Risto Ilmoniemi, for his valuable comments on scientific writing.

Seeing professionals at work, be it engineers or clinicians, has been a great source of inspiration and motivation for me. Special thanks to professor Bryan Young from London Health Sciences, Ontario, Canada, for providing us data and sharing his expertise.

Thanks to my friends for not letting me delve too deep into the work. Finally, I want to thank my family, Kari, Mari, Heikki, and Maria for always supporting my endeavours and encouraging me.

Helsinki, 6.9.2011

Antti E. J. Tanner

Contents

Symbols and abbreviations

Symbols

α	EEG band 8–13 Hz
β	EEG band 13–30 Hz
γ	EEG band above 30 Hz
δ	EEG band 0–4 Hz
θ	EEG band 4–8 Hz
ν	The angle between the center-of-mass and the feature vector outside the convex hull
τ	Time variable in wavelet transform
Φ	Wavelet function
ω	Angular frequency
$\vec{a}$	The average of the past absolute steps
d	Distance
f	Generic function
$\hat{f}$	The Hilbert transform of f
$\vec{F}$	Feature vector
P	Signal power
$\vec{p}$	Weighting vector
s	Scale variable in wavelet transform
t	Time
v_1	Specific set-up
v_2	Sensitive set-up
v_3	Compromise set-up
$x[t]$	Discrete signal
$x(t)$	Continuous signal
x^a	Approximation signal
x^d	Detail signal
z	Analytical function

Abbreviations

AED	Anti-epileptic drug
cEEG	Continuous EEG
CHS	Steps outside convex hull
CNS	Central nervous system
ECG	Electrocardiogram
ECoG	Electrocorticogram
EEG	Electroencephalogram
EMG	Electromyogram
EMU	Epilepsy monitoring unit
EOG	Electro-oculogram
FFT	Fast Fourier transform
FP	False positive
FPR	False positive rate
$\mathrm{FPR_{pa}}$	Patient average FPR
$\mathrm{FPR_{w}}$	Average FPR, weighted with recording durations
GPED	Generalized periodic epileptiform discharge
IF	Instantaneous frequency
ICU	Intensive care unit
$LOGPOW$	EEG power feature
NCSE	Non-convulsive status epilepticus
NCSz	Non-convulsive seizure
PL30	Path length computed from 30 s window
PL90	Path length computed from 90 s window
PL180	Path length computed from 180 s window
PLAVE	Average of PL30, PL90, and PL130
PLED	Periodic lateralized epileptiform discharge
PLW	Weighted path length
RAM	Random access memory
RWL	Long-window (180 s) adaptive random walk algorithm
RWLF	Long-window fixed-step random walk algorithm
RWS	Number of steps in algorithm II
RWSH	Short-window (90 s) adaptive random walk algorithm
SE	Status epilepticus
Se	Sensitivity
$\mathrm{Se_{Int}}$	Integral sensitivity
$\mathrm{Se_{AO}}$	Any-overlap sensitivity
$\mathrm{Se_{AO}}$ 1st	Any-overlap sensitivity to first annotated seizure extended
$\mathrm{Se_{AO}}$ all szs	Any-overlap sensitivity to all seizures
$\mathrm{Se_{AO}}$ pt	Average any-overlap patient sensitivity
$\mathrm{Se_{AO}}$ pt szs	Average $\mathrm{Se_{AO}}$ of all seizure patients
SNR	Signal-to-noise ratio
Sp	Specificity
$\mathrm{Sp_{Int}}$	Integral specificity
$\mathrm{Sp_{Int,\ pa}}$	Average $\mathrm{Sp_{Int}}$
$\mathrm{Sp_{Int,\ w}}$	Average $\mathrm{Sp_{Int}}$, weighted with recording durations
TP	True positive

List of Figures

List of Tables

1 Introduction

The brain works by transmitting electrical signals between neurons. One way to investigate the electrical activity of the brain is to record scalp potential resulting from brain activity. This method is non-invasive; all measurements are made outside the head and no wounds or scars are made. The recorded signal, i.e., potential difference between two positions, is called electroencephalogram (EEG). The word has its origins in Greek: εγκέφαλος (enkephalos) means the brain—or literally "inside the head"—and γράμμα (gramma) means letter or writing.

Recording and investigating signals arising from inside the head has been an active field of research for more than 100 years. However, only during the last 50 years or so, with the breakthrough of digital technology, has EEG become a standard method in medical practice. There are also other methods for monitoring the activity of the brain. Magnetoencephalogram captures the magnetic field caused by electrical activity and functional magnetic resonance imaging can reveal changes in hemodynamics inside the brain. In clinical practice, however, EEG is by far the most common method.

EEG signal can be described by its dominant frequency and power. If the signal has very low power it is called suppressed, or in the extreme case when there is no electrical activity, isoelectric. There is a standard way of attributing Greek letters to different frequency bands. Division of frequencies into these bands was justified by early EEG findings. Nowadays, the most important function of the division is the standardization of EEG vocabulary.

Activity lower than 4 Hz is called delta (δ) activity. Theta (θ) activity is the range of 4–8 Hz, alpha (α) is 8–13 Hz, beta is (β) 13–30 Hz, and activity above 30 Hz is called gamma (γ) activity.

In addition to describing these general features of the signal, neurologists also look for signs of neurological dysfunctions. Interpreting EEG is a very demanding task. Certain artefacts can mimic brain activity and there is often an overwhelming amount of data.

Epileptic seizures form one class of neurological dysfunctions. People with epilepsy have recurrent, unprovoked seizures. However, also people who do not suffer from epilepsy may have seizures [1]. During seizure, there is abnormal electrical activity in the brain. This abnormality is reflected on scalp potentials and hence can be recorded with EEG. Behavioural manifestation can range from subtle finger twitching to convulsions where muscles contract and relax in an uncontrolled fashion, resulting in involuntary body movements. It is also possible that no change in behaviour is seen, or that the change is very subtle. Such seizures are detectable reliably only by monitoring the brain's electrical activity.

Patients treated in intensive care units (ICUs) are critically ill. Because of their critical condition, ICU patients are often artificially ventilated and sedated. This helps them withstand care-giving operations. Critically ill patients tend to have neurological problems, too.

Lately, it has been shown that a considerable amount of ICU patients suffer from seizures. Some seizures are convulsive and can thus be noticed by the bedside staff

and treated accordingly. The majority of seizures encountered in ICUs, however, are non-convulsive. If the unit does not have a protocol to monitor brain activity and to constantly review the data, these seizures will not be detected, or they will be detected when the possibility to intervene has already passed.

Seizures constitute a medical emergency and have to be medicated. If medication is not started promptly, effects on the patient can be detrimental. [2]

There are different types of non-convulsive seizures. We are focusing on seizures that follow a dynamical pattern where signal characteristics change consistently between samples. In other words, we are looking at gradual changes, or evolution, in time-courses of signal features.

This thesis is a part of a larger project where ICU neuromonitoring practices as a whole are updated. The main goal of this thesis is to develop and evaluate algorithms for detecting non-convulsive seizures. More specifically, the developed methods should detect seizures of evolutionary type. Furthermore, seizures that cannot be detected by developed methods should be identified, as well as the main causes for false positive detections.

Sect. 2 provides a brief introduction to EEG, seizures and to the prior art. In Sect. 3, data set and employed mathematical methods are described. Features that are extracted from EEG and fed to detection algorithms are presented in Sect. 4. The detection algorithms developed for the task are described in Sect. 5. The main results of the development work are presented in Sect. 6, and their impact is discussed in Sect. 7.

The author has contributed all material presented from Sect. 4.2 onwards, with the exception of the idea of algorithm I.

2 Background

This thesis is a part of a larger effort to renew brain state monitoring practices in ICUs. A novel electrode cap and an artefact rejection algorithm were being developed in parallel with this thesis. The goal of the project as a whole is to provide the ICU staff an easy-to-use means for monitoring their patients' neurological states. Furthermore, the staff should be able to easily interpret the information provided by the algorithms and gain objective evidence to support decision making.

Using the full conventional EEG measurement set-up is time-consuming [3]. The novel electrode cap is designed to be as easy as possible to put on the patient and to align with anatomical landmarks. It can maintain a good electrical contact throughout prolonged recordings. To facilitate the set-up process and data processing, we are using only a subset of the full electrode montage.

Artefact rejection is a critical part of any biomedical signal processing application. We want to ensure that only high-quality data are passed on after this step. In the ICU, several abnormal EEG patterns are present because the patients are in a critical condition and the environment is anything but calm and controlled. This makes it difficult to design a reliable artefact rejection algorithm. We use additional information from accelerometers integrated to the electrode cap to detect motion artefacts accurately [4].

At the heart of the concept is the signal processing algorithm. The algorithm should provide an accurate picture of the patient's state and address the problem of non-convulsive seizures that are nowadays mostly undetected and, therefore, untreated. This thesis focuses on the development of the EEG signal processing algorithm. The rest of this section is devoted to familiarizing the reader with the environment and the techniques used.

2.1 Basics of EEG

First human EEG was recorded by Hans Berger in the 1920s. As a pioneer in the field, it was he who coined the term electroencephalogram. His work was based on initial animal brain function studies performed by Richard Caton in the 19th century. Caton and Berger described several normal and abnormal EEG patterns, among them α and β waves. [5]

In the field of seizure detection, the earliest studies were performed in the 1930s when Fisher and Lowenback described epileptiform spikes [6]. Throughout the latter half of the 20th century, with the dawn of digital recording techniques and widely accessible computing power, we have seen a surge of studies that describe both the origins of EEG and what clinicians and researchers can infer from it. Today, EEG is widely accepted as a standard measurement technique.

A modern EEG device consists of a set of electrodes, an amplifier, a data storage unit, and a display unit. The electrodes are fixed to the subject's scalp and conductive gel is applied to the electrode–scalp interface. Electrode positions can be chosen according to the standardized 10–20 system [7] (see Figs. 1 and 2) or by prior information about patient's etiology. In some cases, usability aspects might favour

setting the electrodes on hairless area. Sub-hairline montage is a commonly used way to facilitate electrode set-up. In this set-up, electrodes are placed on temples and on forehead, near to the subject's hairline.

An EEG channel, or derivation, at its simplest form consists of two electrodes that are connected to an amplifier. The output signal is the amplified potential difference between the electrodes. In the common reference mode, one electrode is used as a reference for all other electrodes. A ground electrode can be used to reduce mains interference. A collection of derivations is called a montage.

EEG practitioners are mainly interested in signals generated by brain activity. Neurons communicate by releasing neurotransmitters in synaptic clefts. Neurotransmitters can selectively open and block receptors, causing a flux of ions from the synaptic cleft to the post-synaptic cell or vice versa. The flux is driven by the difference in intra-cellular and extra-cellular ion concentrations. The flux of ions forms a current dipole. The amount of current resulting from a single synapse is not enough to generate a measurable scalp potential. When a group of neurons is activated simultaneously, the net effect can be measured on the scalp. Action potentials do not produce easily measurable potentials at the scalp. [9]

EEG monitoring can provide information about brain activity. However, solving

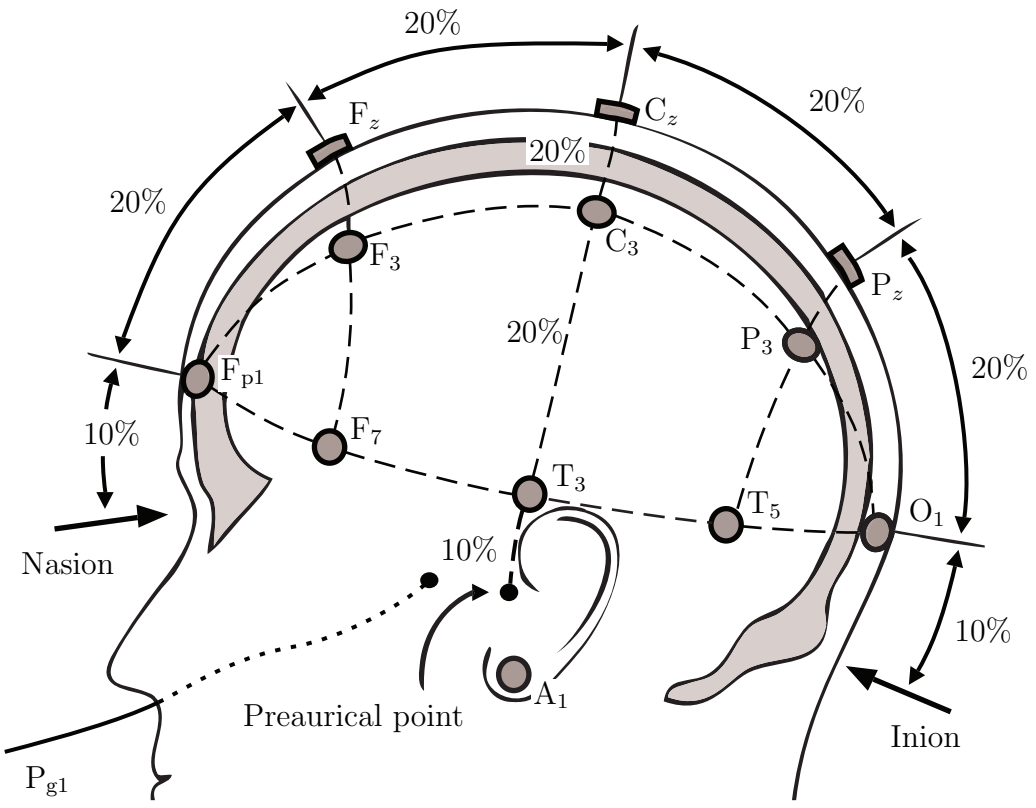

Figure 1: A sagittal view of electrode placement according to the international 10–20 system. Modified from the original figure by Malmivuo et al. [8].

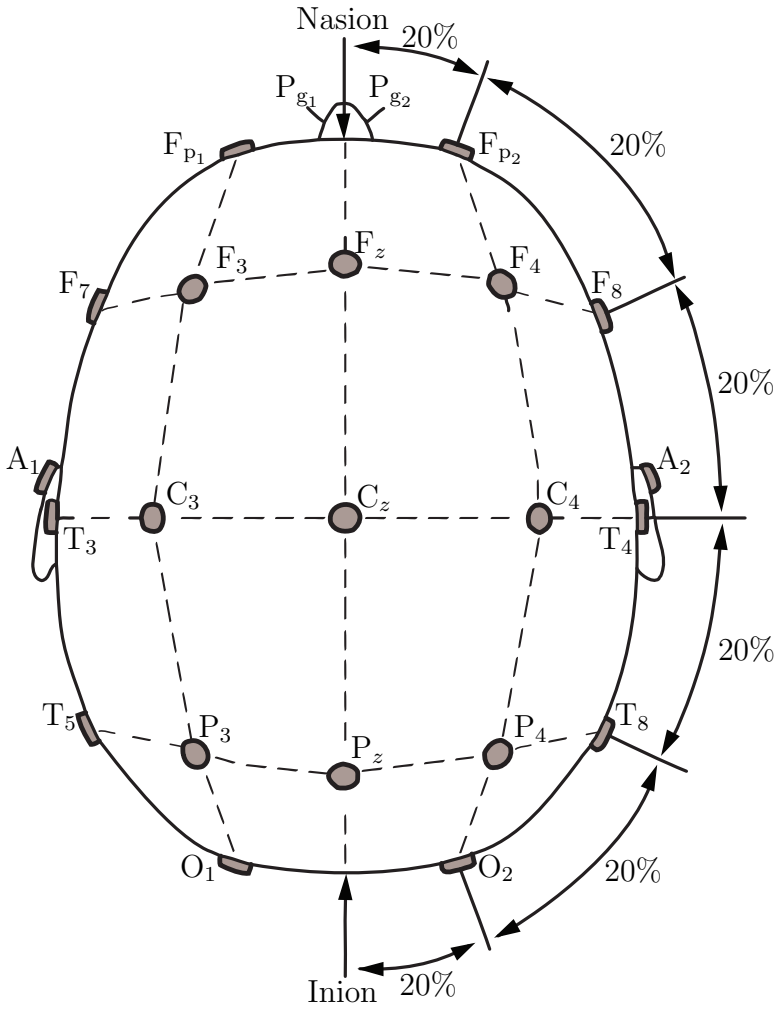

Figure 2: An axial view of electrode placement according to the international 10–20 system. Modified from the original figure by Malmivuo et al. [8].

the inverse problem, i.e., solving the source dipole distribution given the potential distribution at scalp, is a difficult task. Without *a priori* assumptions about the source dipole distribution and regularization, the problem is not even well-defined. In this thesis, we take the EEG signal as a real-valued time course *per se*, without considering the actual electrical sources of the signal. The sources might provide interesting information but hardly contribute to the algorithm under development. Fig. 3 shows an example of adult EEG recorded in the ICU.

The use of EEG as a diagnostic tool has been hindered partly by the lack of common vocabulary among practitioners. Lately, such a nomenclature has been proposed by the American Association of Neurologists [10]. We will use that naming convention in this thesis. For classifying EEG recordings, we follow the classification scheme for comatose EEG proposed by Young et al. [11]. The scheme is presented in Table 1.

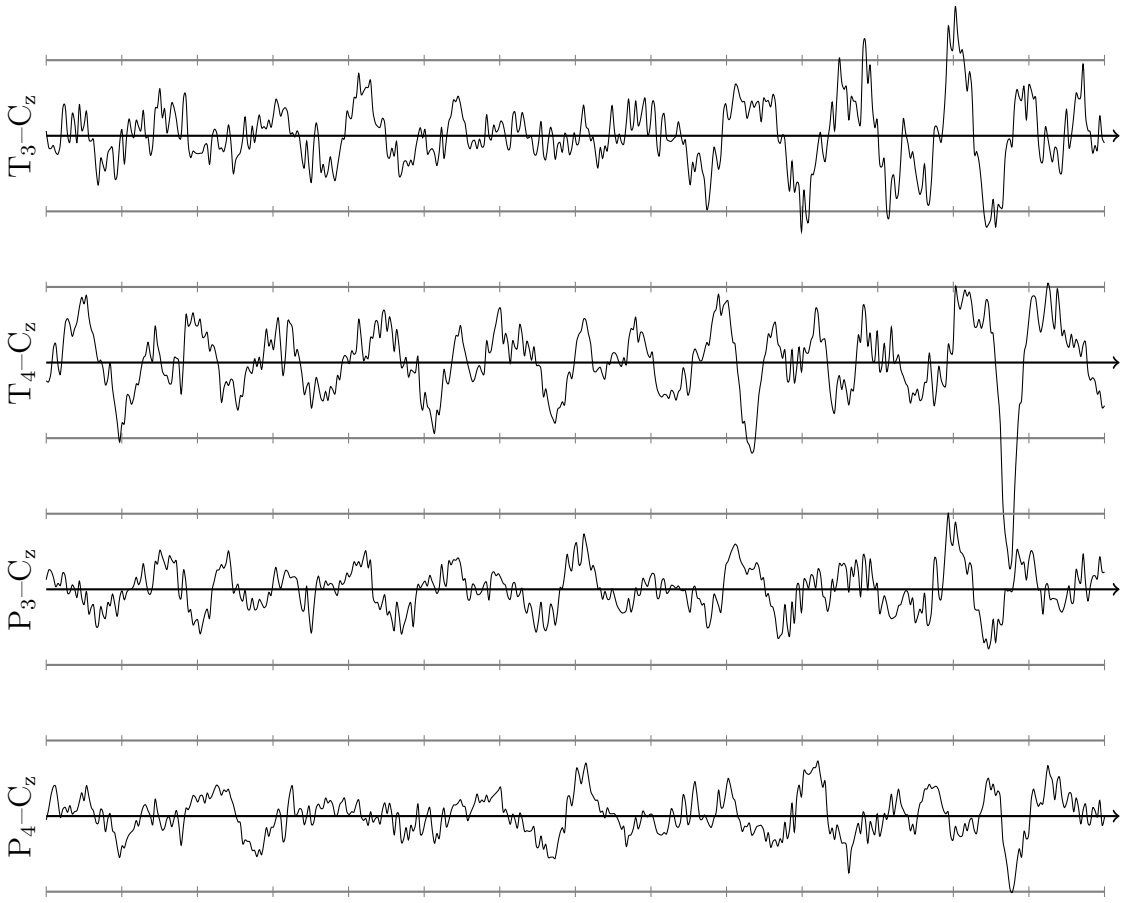

Figure 3: Normal adult EEG with reactivity and variability (category IA). Recorded in the ICU. The black line marks baseline, grey lines show $\pm 50\,\mu V$ values, and tick marks are printed every second. The names of the derivations are indicated on the left side of each signal.

2.1.1 Common artefacts

The goal of EEG monitoring is to infer the neurological state of the patient. Artefacts can contaminate the measurement and even lead to a false diagnosis. Artefacts can be divided into physiological and mechanical artefacts. Electromyogram (EMG), electrooculogram (EOG), electrocardiogram (ECG), ballistic effect, and glossokinetic potential are examples of physiological artefacts. EMG is caused by muscle activity and is typically manifested in high-frequency frontally predominant activity. EOG reflects the movement of eyeballs. Since there is a voltage between the cornea and the retina, the eyeball acts like a dipole, also contributing to the scalp potential. EOG is best seen in frontal electrodes. ECG is caused by electrical activity of the heart. The ballistic effect overlays a pulse-synchronized signal on EEG due to pulsation of blood.

Typical mechanical artefacts are mains interference, mechanical movement of the electrodes and bed vibrations. Moving electrode leads also cause an artefact.

Table 1: Comatose EEG classification scheme [11].

Category		Subcategory
I	δ or $\theta > 50\%$ of record (not θ coma)	A. With reactivity B. Without reactivity
II	Triphasic waves	
III	Burst–suppression	A. With epileptiform activity B. Without epileptiform activity
IV	α/θ/spindle coma (unreactive)	
V	Epileptiform activity (not in burst–suppression pattern)	A. Generalized B. Focal/multifocal
VI	Suppression	A. 10 µV to 20 µV peak-to-peak B. <10 µV peak-to-peak

This is evident if we consider the lead as a conductive rod that is moving in an external field. Bending the leads can cause artefacts through the triboelectric effect. Hirsch and Brenner have documented exemplary artefacts [10]. All these artefacts are present in the ICU and extreme care must be taken to ensure that no wrong decisions are made because of compromised signal quality.

We must bear in mind, however, that a signal component that we might consider artefactual in the current application, might actually be the most interesting part of the signal in another study. For example, if EEG of a sedated patient shows increasingly high amount of EMG, we can conclude that the level of anaesthesia might need adjustment [12].

2.1.2 Special EEG techniques

Special techniques complement the normal EEG recording scheme. Continuous EEG (cEEG) means recording EEG continuously for extended periods of time. This leads to challenges in electrode design and in data storage methods. The electrodes should be designed so that their impedance levels do not deteriorate too much during long recordings. Patient safety also becomes a concern as skin contact is maintained for several hours or even days. Lengthy measurements result in huge amounts of data that need to be processed and stored. However, evidence suggests that cEEG is necessary to provide an accurate picture of the patient's state [13–17].

In video EEG, a camera is used for recording video simultaneously with the EEG. Video can help clinicians identify artefacts. It provides behavioural correlates with EEG. Using depth electrodes to record electrocorticogram (ECoG) can provide a less distorted and artefact-free signal and thus facilitate interpretation [10, 18]. ECoG is, however, an invasive modality.

Analysing hours of EEG data is extremely time-consuming. It would be advantageous to be able to compress the data and present only the relevant epochs or

trends to clinicians. Agarwal and Gotman have presented one example for summarizing EEG data [19]. Even though such semi-automatic post-processing methods have been designed, the fact remains that raw EEG signal must be available for later review.

2.2 Seizures

Epilepsy is a term used for a group of neurological disorders. Individuals with a diagnosis of epilepsy have recurrent, unprovoked seizures [1]. A seizure is an abnormal electrical discharge in the brain [20]. Some people have genetic features that elevate the risk of developing epilepsy. Epilepsy can also emerge due to a structural abnormality. Idiopathic epilepsy means that the cause of the disorder is unknown. Types of seizures, their intensity, frequency, and duration vary a lot between patients, but it is common that the same pattern is repeated on each occasion on a given patient. About 0.6% of the general population suffers from epilepsy. The prevalence varies, however, between age groups. [21, 22]

Seizures can also occur in individuals without diagnosed epilepsy. In contrast to epileptics, these seizures are not recurrent or they are provoked. ICU is an example of an environment where seizures have been encountered in patients without a diagnosis of epilepsy. Typical etiologies for patients whose first seizures are encountered in the ICU are summarized in Table 2. All of the studies referenced here have listed hemorrhages in the head as a common etiology.

Seizures can be divided into two groups according to behavioural correlates of the electrographic activity. In convulsive seizures the patient has visible convulsions, e.g., rhythmic jerking. If there are no visible changes, or if they are subtle, such as nystagmus, eye deviation, or myoclonus, the seizure is called non-convulsive. For non-convulsive seizures, cEEG remains the best available detection method. [14, 16]

According to a generally accepted definition, when there is continuous or nearly continuous seizure activity for a minimum of 30 min, status epilepticus (SE) is diagnosed. SE is a medical emergency that requires intervention [2]. The corresponding term for persistent non-convulsive seizures is non-convulsive status epilepticus (NCSE). Seizures that emerge from background EEG are called isolated seizures. Cyclic seizures show a pattern with seizures starting at almost constant intervals.

Prevalence estimates of seizures in the ICU are summarized in Table 3. Even though there is a large variation in the numbers reported, it can be concluded that seizures in the ICU are much more common than prevalence of epilepsy in the general population would suggest. Strikingly many seizure patients have only non-convulsive seizures.

Because cEEG monitoring is not a standard procedure, it is hard to estimate the general prevalence of seizures in the critically ill. In retrospective studies, data that was recorded before is reviewed and findings are reported. The fact that the EEG was recorded in the first place means that there was an indication for doing so. Thus, such studies may have selection bias. On the other hand, if a prospective study is targeting only a specific group of patients, we cannot draw conclusions about the general prevalence of the studied phenomenon.

Table 2: Common etiologies of ICU patients with seizures.

Study	**Etiologies**
Young et al. [23] (Non-convulsive seizures)	Multiple organ failure, anoxic-ischemic encephalopathy, subarachnoid or intracerebral hemorrhage, prior seizures
Claassen et al. [16]	Prior seizures, CNS infection brain tumour, previous neurosurgical intervention, subarachnoid hemorrhage, decrease in the level of consciousness
Alroughani et al. [24] (NCSE)	Hypoxic-anoxic injury, intracerebral hemorrhage, stroke

CNS = Central nervous system.

There are two reasons why the prevalence of seizures in the ICU is elevated. First, the patients are presented with a variety of therapeutic drugs that may lower seizure threshold. Second, because the patients are critically ill, with possible multi-organ dysfunctions, there are plenty of possible causes for cerebral disturbances. [25]

Seizures are treated using anti-epileptic drugs (AEDs). Standard treatment of prolonged seizure activity consists of airway maintenance, oxygen, and intravenous

Table 3: Prevalence of seizures in ICU. Modified from a publication by Friedman et al. [14]. Included only cEEG studies.

Study	**N**	**Percentage of patients with seizures**	**Percentage of seizure patients with only NCSz**	**Design**
Jordan [26]	124	35	74	Ret.
DeLorenzo et al. [27]	164	48	100	Pros.
Vespa et al. [28]	94	22	52	Ret.
Vespa et al. [29]	109	19	79	Pros.
Claassen et al. [16]	570	19	92	Ret.
Pandian et al. [30]	105	68	N/A (27% NCSE)	Ret.
Jette et al. [31]	117	44	75	Ret.
Claassen et al. [32]	102	31	58	Ret.
Oddo et al. [33]	201	10	67	Ret.
Alroughani et al. [24]	451	Overall 9.3% NCSE		Ret.

Ret. = retrospective, Pros. = prospective

diazepam. During medication, the cause of seizures should be investigated. If the condition worsens, intravenous anaesthesia, intubation, and ventilation are required. Thiopentone or propofol can be titrated until a burst–suppression pattern is seen in EEG. [2]

Accumulating evidence suggests that also non-convulsive seizures and seizure activity that does not qualify as SE should be treated. A widely supported view is to consider seizure activity longer than 5–10 min as a condition that requires intervention [15]. Periodic epileptiform discharges (PLEDs) might not count as a seizure, but initiating prophylaxis has been suggested as a reasonable action if such patterns are found [34].

For non-convulsive seizures, seizure duration and delay to diagnosis have been found to be associated with increased mortality. However, it should be noted that this patient population consists of critically ill. It might be practically impossible to tell whether the mortalities are caused by the original underlying etiologies or by the neurological dysfunctions resulting from them. [14, 23]

EEG is the standard tool in seizure studies. Once EEG has been recorded, a seasoned expert should give a report on the findings. Annotating seizures retrospectively is a very demanding task. There is a considerable disagreement even among those skilled in the art as to where the begin and end annotations should be placed [35]. To provide a concrete means for seizure classification, a scheme for what should be called a seizure has been proposed (see Table 4). Our development work is based on these criteria.

Table 4: Criteria for seizure detection [23].

Guideline: To qualify at least one of primary criteria and one or more of secondary criteria, with discharges for more than 10 s
Primary criteria:
1. Repetitive generalized or focal spikes, sharp waves, spike-and-wave or sharp-and-slow wave complexes at $>3\,\mathrm{s}^{-1}$.
2. Criterion 1 at $<3\,\mathrm{s}^{-1}$ and secondary criterion 4.
3. Sequential rhythmic waves and secondary criteria 1, 2, and 3 with or without 4.
Secondary criteria:
1. Incrementing onset: increase in voltage and/or increase or slowing of frequency
2. Decrementing offset: decrease in voltage or frequency
3. Post-discharge slowing or voltage attenuation.
4. Significant improvement in clinical state or baseline EEG after AED

2.2.1 Epileptiform EEG

The adjective epileptiform means something that is related to epilepsy [36]. Epileptiform EEG thus refers to EEG patterns that are often found in epileptics. Epileptiform patterns are abnormal, but as such they do not constitute a seizure. Even a patient without seizures can have some epileptiform patterns present in the EEG. EEG during seizure is called ictal and between seizures inter-ictal.

This subsection summarizes some common EEG findings in the critically ill. More examples on how to interpret EEG findings in the critically ill are reported by Hirsch and Brenner [10] and by Chong and Hirsch [34].

The first described indication of epileptiform activity was the presence of spikes. An example of EEG with several spikes is shown in Fig. 4. A spike is a sharply contoured waveform with a duration of 20–70 ms [9]. Spikes are longer in duration than EMG activity. If spike rate exceeds $3\,\mathrm{s}^{-1}$ continuously for more than 10 s, the EEG can be classified ictal [23]. Classification as ictal is warranted also if spike rate is lower and there is a response to administrated AED (see Table 4).

Epileptiform discharges constitute a severe EEG finding. Fig. 5 shows an example of generalized periodic epileptiform discharges (GPEDs). Generalized activity is present on several channels on both hemispheres. Focal epileptiform activity is visible only on a few derivations.

A very common EEG finding in the ICU is the burst–suppression pattern. An example is shown in Fig. 6. This finding is abnormal but is not always related to a neurological dysfunction. The burst–suppression pattern emerges also when the patient is under heavy medication with sedatives. If medication is further increased, the EEG would ultimately become totally suppressed, or isoelectric. The burst-suppression pattern is often described by the amount of suppressed EEG in the epoch. The example epoch has a burst–suppression ratio of about 75%. It should be noted that the bursts may contain epileptiform discharges.

When reviewing an EEG recording with seizures, one can often notice a pattern that evolves from the initial, only barely noticeable epileptiform patterns to a seizure with periodic discharges followed by post-ictal suppression. It must be stressed, however, that not all seizures follow the same pattern. Fig. 7 shows an example of a seizure with quite clear on-set and very clear off-set. Even the unskilled in the art can easily follow how signal frequency and amplitude evolve during the event.

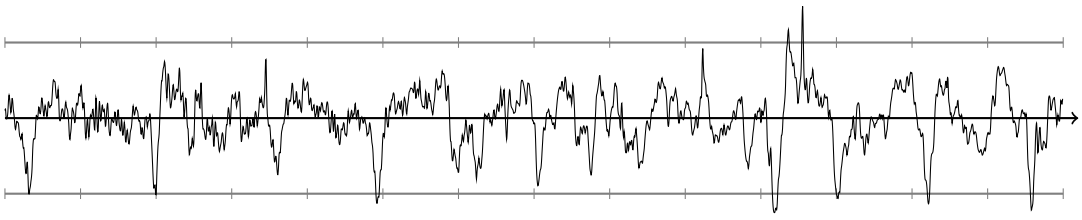

Figure 4: Spikes. The black line marks baseline, grey lines show $\pm 50\,\mu\mathrm{V}$ values, and tick marks are printed every second.

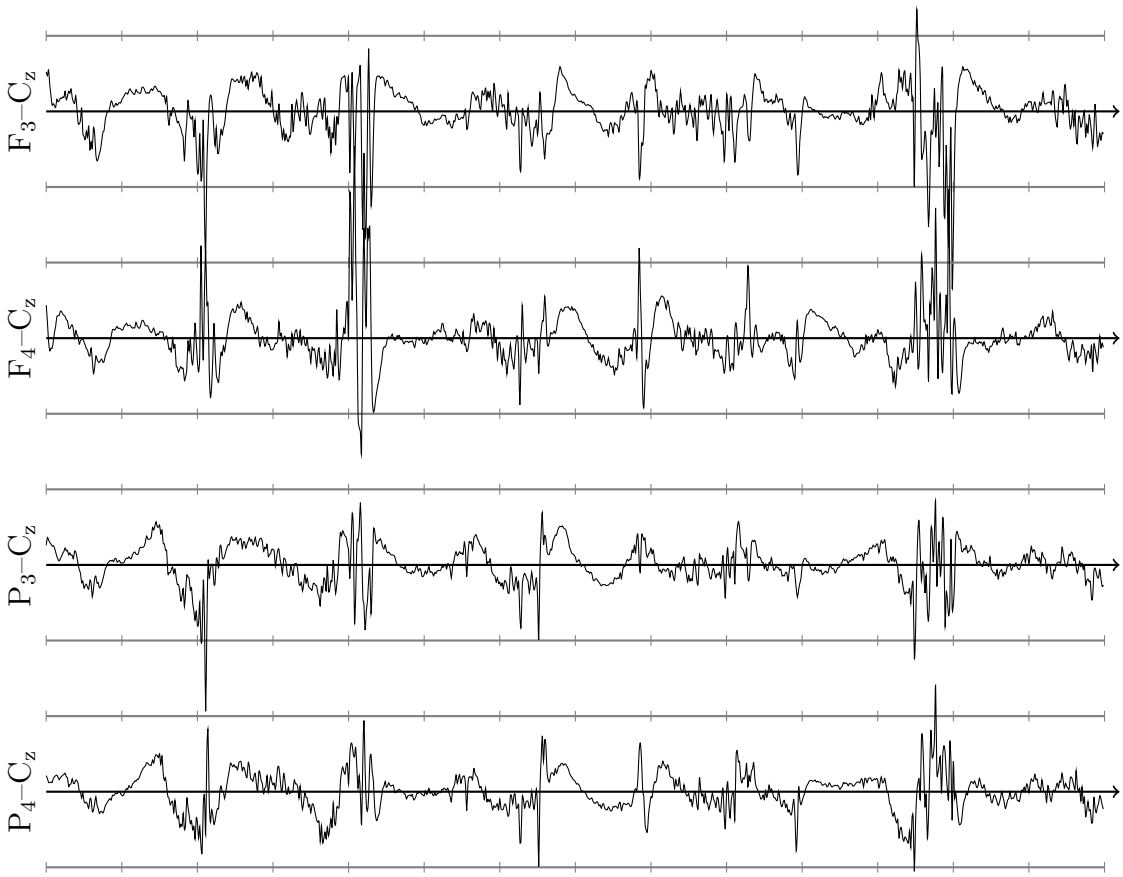

Figure 5: GPEDs (generalized periodic epileptiform discharges). The black line marks baseline, grey lines show ±50 µV values, and tick marks are printed every second.

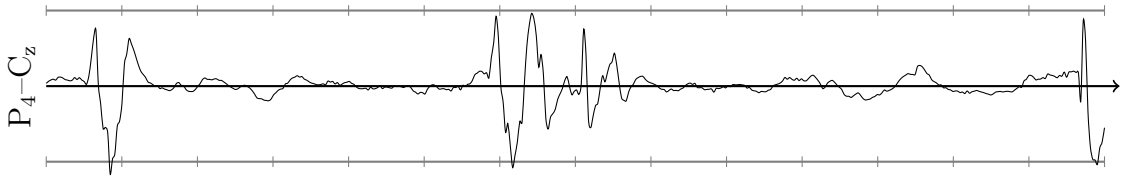

Figure 6: An example of a burst–suppression pattern. The black line marks baseline, grey lines show ±50 µV values, and tick marks are printed every second.

2.3 Development background

2.3.1 Neurological monitoring in ICUs

The motivation to use on-line neuromonitoring in ICU arises from several reports describing the prevalence of non-convulsive seizures in critical care (see Table 3). In many places, the current protocol allows EEG monitoring only if there is a doubt of a neurological problem. In this case, an EEG technologist brings a portable EEG

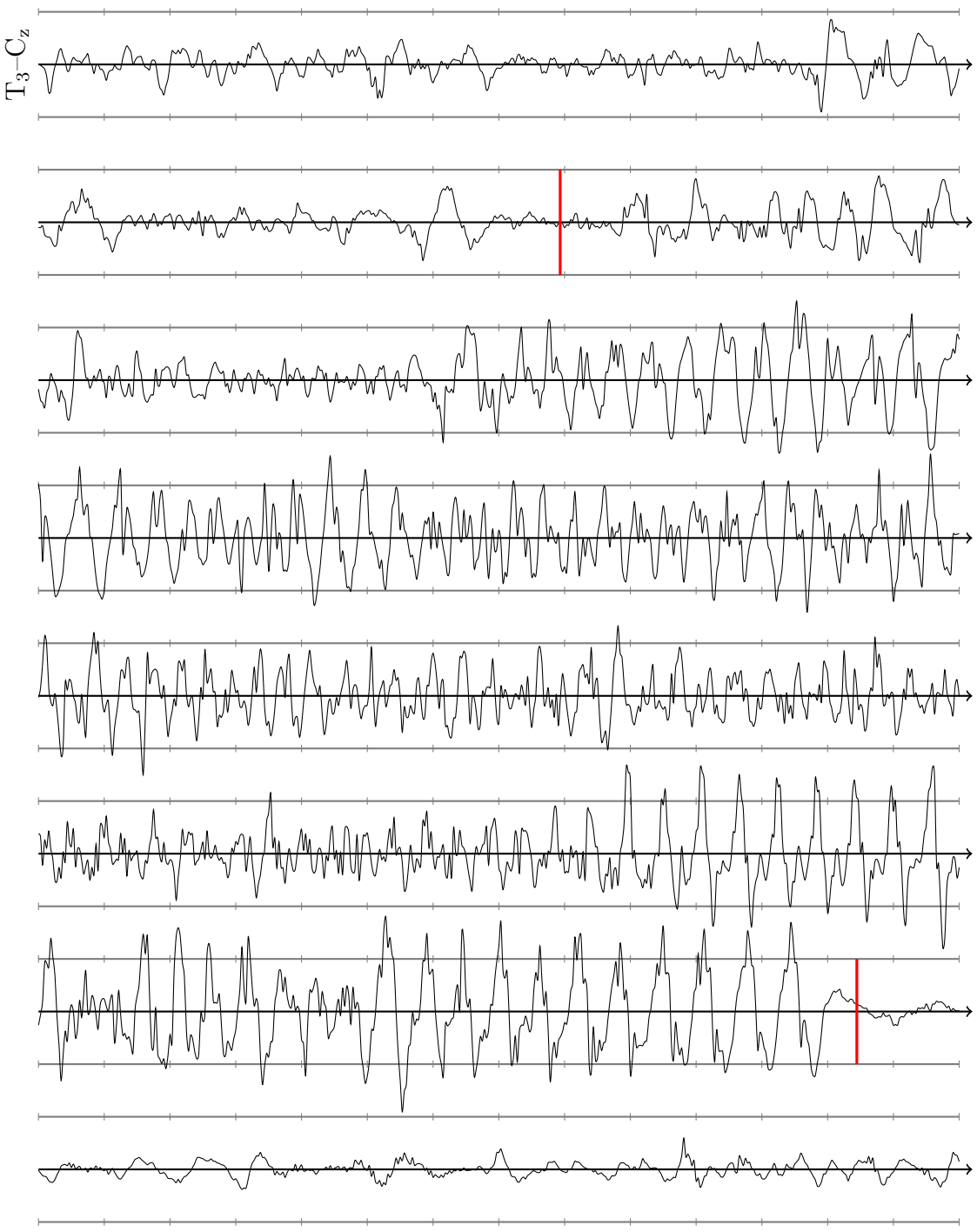

Figure 7: An example of a seizure recorded in ICU. Seizure on-set and off-set, as annotated by a clinician, are marked with vertical red bars. The same seizure is also used later on as an example when describing the developed algorithms. This seizure follows a pattern typical for evolution seizures: increasing frequency followed by an increase in amplitude and, finally, by post-ictal suppression. The black line marks baseline, grey lines show $\pm 50\,\mu V$ values, and tick marks are printed every second.

system to the ICU, connects the electrodes and records for approximately 30 min. The data are later reviewed by a neurologist who decides which actions should be taken. There are typically 1–2 assessments per day in a non-neurological ICU [14].

This scheme is inefficient for two reasons. First, there is no guarantee that a seizure appears during those 30 min. In a retrospective study, it was found that only 56% of seizure patients had their first seizure during the first hour of the cEEG recording. After 48 h, 93% of the patients had encountered their first seizure [16]. Second, in the current scheme, there is a considerable delay between the recording and the possible intervention.

2.3.2 Prior art

Several algorithms have already been designed for seizure detection in epilepsy monitoring units (EMUs) and for neonatal patients. ICU, however, has so far been out of their scope. Gotman, a recognized researcher in the field, has reviewed the general principles of seizure detection [37, 38].

Many different approaches have been experimented with in order to produce EEG features that are specific to seizures and yet sensitive and generic enough to capture the majority of them. A selection of seizure studies is summarized in Table 5. Seizure detection has been approached as a machine-learning problem consisting of two main steps: generating features and designing a classifier.

The first methods in the field did not actually aim at seizure detection, but at compressing data and highlighting events for neurologist's later review. These semi-automatic detection systems can speed up the reviewing process, but they impose a delay to intervention. There are several reports describing widely used compression

Table 5: Summary of some published seizure detection algorithms.

Study	**Used features**	**Classifier**
Agarwal et al. [19]	A, dominant frequency, energy	K-means
Gabor et al. [39]	Frequency, time course	SOM
Firpi [40]	Wavelet & FFT	RBF
Zandi et al. [41]	Rhythmicity & energy	LDA
Tezel et al. [42]	Time course descriptors	ANNAAF
Guo et al. [43]	Wavelet line length	ANN
Zandi et al. [44]	Rhythmicity, consistency, relative energy	CUSUM

A = amplitude, SOM = Self-organizing map,
FFT = Fast Fourier transform, ANN = Artificial neural network,
RBF = Radial basis function ANN,
ANNAAF = ANN with adaptive activation function,
LDA = Linear discriminant analysis, CUSUM = Cumulative sum control chart

tools such as compressed spectral array, density spectral array, spectrogram and non-linear energy operator [17, 19, 45, 46].

In the past, the most widely applied tactics was to analyse the frequency content of the signal by applying some variant of the Fourier transform. When using this method, there is a considerable trade-off between time and frequency resolutions. If spectrum is evaluated in short windows, the time resolution is fair, but the frequency resolution is poor. If a longer window is used, the frequency resolution increases but the information is less concentrated in time.

A more modern approach is to use wavelet transform. This method is more thoroughly presented in Sect. 3.2.1. Wavelet decomposition allows representing signal's properties in different scales. This way both the time and the frequency content can be assessed with a relatively good resolution.

Some features derivable directly from the raw EEG have also been experimented with. The most common features are signal power and zero crossings. They have a very low computational complexity, which is an advantage when designing an on-line algorithm.

When the features have been generated, it remains to devise the decision making method. This is a typical problem of supervised learning. Given a development data set with known desired outcomes, one should design a system that performs well in the development data. Furthermore, the classifier should be able to generalize and show good performance also in previously unseen evaluation data set. Annotations made by a neurologist are often considered as the ground truth in the problem setting.

Machine-learning methods that have been applied in this field include expert systems, decision trees, clustering algorithms, self-organizing maps, and a variety of artificial neural network configurations. While the sophisticated machine learning techniques can enhance the performance of the algorithm, they can also be cumbersome for the end-user to interpret. The simpler the system is, the easier it is for specialists to learn to trust and to understand it.

Some seizure detection algorithms have been made available for EMUs and for neonatal patients. We present here the most widely used systems.

Gotman has been involved in the field of EEG monitoring and seizure detection since the 1970s. His algorithms are distributed by Stellate. First algorithms used a decomposition of EEG into elementary waves and inspected their properties [47, 48]. Another module was added to exclude common causes for false positive detections [49].

Persyst Development Corporation also offers seizure detection software. The algorithm, Reveal Rosetta, is also promised to hold potential for ICU use. The structure of the algorithm is largely unpublished. [50]

CNET is a non-commercially distributed algorithm for seizure detection. It uses cepstral features to describe EEG. [39]

2.3.3 Design drivers for the new algorithm

The proposed monitoring scheme is summarized in Fig. 8. This thesis focuses on the evolution indicator branch. According to previous work done in the development team, it has been established that power of the wavelet transformed signal together with instantaneous frequency (see Sect. 3.2) could serve as a good starting point for the algorithm development.

Accumulating evidence shows that signal power and frequency content contain information relevant for seizure detection. These features have been an integral part in many studies (see Table 5) and they are also appreciated in the seizure criteria (see Table 4). Primary criteria 1 and 2 remain out of the scope of these features, but seizures that satisfy the third criterion should be detected. We can also conclude that static variables or trends do not give the full picture of the patient's state. Instead, one must look for certain kind of dynamics in the feature values, or evolution. In this thesis, we will use the term evolution to refer to continuous, consistent changes in feature values. This definition excludes sudden jumps and changes that mostly cancel out. Seizures with such characteristics are called evolution seizures in the context of this work. Finally, it is desirable to keep the features and decision making process as simple and tangible as possible.

Based on interviews with ICU doctors, we have established certain goals for the development project. Naturally, the system should be reliable in terms of specificity

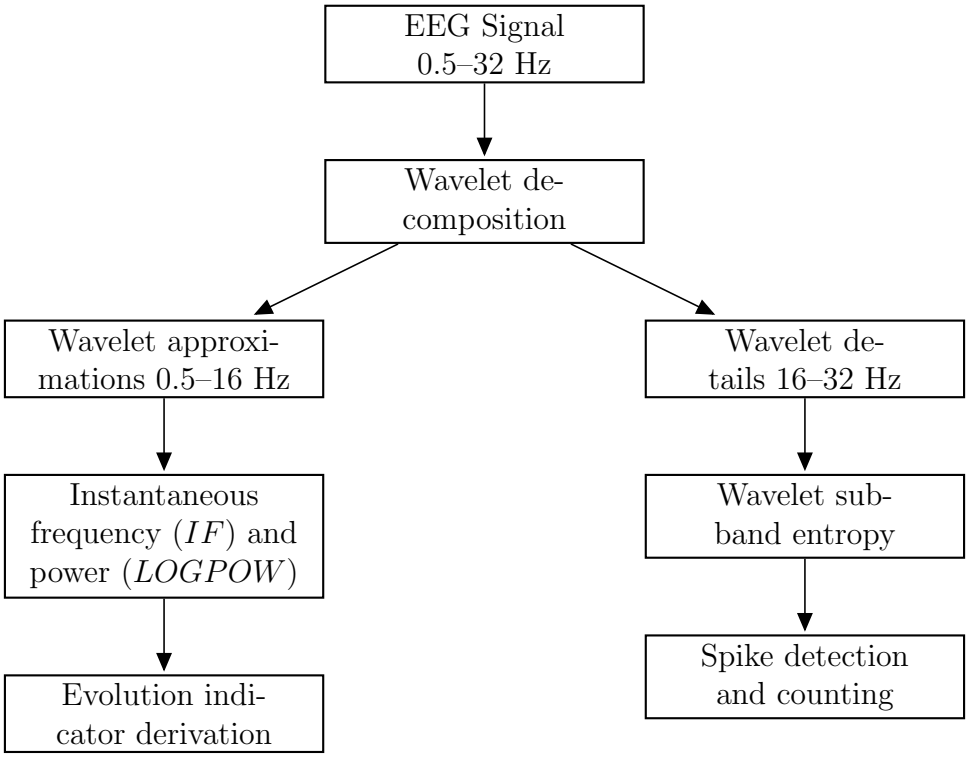

Figure 8: Schematic overview of the seizure detection algorithm. Särkelä has published a description of the wavelet subband entropy [51].

and sensitivity. It must be able to detect both isolated seizures and prolonged ictal activity. Ability to compress data and to present long-term trends that monitor how the patient's state has progressed would be a desirable feature. Finally, the system should be user-friendly both in terms of visualization and in using the electrode cap.

The project aims at developing monitoring software that presents relevant information at the bedside and at the event of seizure activity alerts both the bedside staff and a neurologist. A summary containing the raw EEG could be sent to the neurologist who has remote access to the information system. After administering AED, the effect could be followed by both the neurologist and the bedside staff on real time. The bedside staff cannot be constantly paying attention to the monitor, and neurologists do not want any unnecessary disturbance caused by irrelevant events. For these reasons, it is important to minimize false positive alerts.

By interviewing experts in the field, we have established the following goals for the seizure detection algorithm:

- Every patient with seizures should be detected
- On each seizure patient, we should reach 80–90% sensitivity
- An acceptable rate of false positive detections is about 1 in 8 h

These are the ultimate goals of the project. At this stage, however, we can relax the specifications since we are designing only one part of the final method. Spike detection, for example, is an integral part of the final method but is not included in the analyses conducted in this thesis.

3 Materials and methods

3.1 Data set

The data for this study have been collected at London Health Sciences Centre, Ontario, Canada. The patients included in the study were receiving critical care.

Data were collected using two different devices. A sub-hairline EEG was recorded by a device manufactured by Datex-Ohmeda. Standard scalp EEG was recorded by a device manufactured by XL-Tek. Data from both recordings were analysed retrospectively by an expert neurologist, Dr. G. Bryan Young. Two devices were used to study how sensitive and specific the limited-coverage sub-hairline montage is compared to the standard 10–20 system.

Because of recent reports implying that sub-hairline EEG has low sensitivity for detecting seizures [14, 52, 53], we will use normal scalp EEG recordings as the development data. Unfortunately, in some recordings, periods of data had been removed before the data were delivered. For our use, we extracted from the original data files derivations F_3–C_z, F_4–C_z, T_3–C_z, T_4–C_z, P_3–C_z, and P_4–C_z.

Each recording was categorized according to the established method (see Table 1). In addition, electrographic seizures were annotated by the expert. Results of the data classification are presented in Table 6. Only ten first seizures for each recording were annotated. Statistics of annotated seizures are given in Table 7.

Table 6: Findings in expert's review of the data set. Classes I A and I B have been collapsed into class I.

Young's coma classification	**N**	**Percentage**
I	67	37.4%
II	22	12.3%
III A	11	6.1%
III B	15	8.4%
IV	7	3.9%
V A	16	8.9%
V B	23	12.8%
VI A	19	10.6%
VI B	9	5.0%
Without classification: 6		
Patients belonging to two classes: 16		
Total patients: 179		
Total recordings: 260 (>150 d of data)		

Table 7: Summary of seizure patients in the development data set. Durations refer to annotated seizures. For each recording, only ten first seizures were annotated. Sub-recordings from each subject have been pooled together for this table.

Identifier	**Min [s]**	**Max [s]**	**Median [s]**	**N**
case016	19	27	21	3
case035	34	399	111	13
case098	28	32	29	3
case101	9	36 h	357	25
case102	68	68	68	1
case103	26	45 min	165	10
case104	9	105	19	4
case106	35	35	35	1
case109	88	88	88	1
case119	118	118	118	1
case122	41	54	42	3
case126	14	84	51	10
case130	15	15	15	1
case144	71	162	93	11
case156	66	66	66	1
case157	10	72	42	31
case159	48	94	61	3
case177	9	24 h	20	12
case178	8	79	48	5
Recordings with seizures: 28				
Seizure patients: 19				
Annotated seizures: 139				

3.2 Mathematical methods

We will be working with a two-dimensional feature space. One feature captures the instantaneous frequency of the signal and the other one measures signal power. Signal power is computed from the stationary wavelet transform. For instantaneous frequency computation, we apply the Hilbert transform. The transforms are briefly explained below. For the interested, there are several thorough guides available [54–56].

3.2.1 Wavelet transform

Traditional signal analysis methods rely on the Fourier transform and its variants. In Fourier analysis, the signal is transformed to a new basis spanned by harmonic functions. Taking inner products with the basis functions, we find out how much of the signal is explained by each basis function. In the case of a harmonic basis, we can immediately recognize which frequencies are present in the signal.

A drawback with harmonic basis functions is their lack of time resolution. Because the basis functions are not localized in time, we cannot pinpoint how and when the frequency content changes within the computation window. Sliding windows, different windowing functions, and the short-time variant of the Fourier transform have solved some of the issues.

Wavelet analysis approaches signal decomposition from another point of view. Instead of non-localized harmonic functions, wavelet analysis makes use of square-integrable, localized basis functions. The basis is formed by translations and dilations of one basis-generating function called the mother wavelet.

The following discussion is mainly based on the documentation of the software package used in this study [57]. The first introduced mother wavelet was the Haar wavelet (see Fig. 9a). It is straightforward to show that it has the properties required for a basis. In many applications, however, the Haar wavelet is not the optimal choice. A simple example is representing a signal with a sloping line. We would need a lot of Haar wavelets to represent such a simple signal.

An example of a more advanced mother wavelet and its transformation is shown in Fig. 9b–c. From the graphs, it is not immediately clear that these functions, the Daubechies-5 wavelets, constitute a basis. Daubechies wavelets, named after the Belgian mathematician and physicist Ingrid Daubechies, are nowadays perhaps the most widely used family of wavelets.

Let $\Phi(s, \tau)$ be the mother wavelet. Basis functions are generated by dilating the mother wavelet, or in other words, altering its scale s. Translations are achieved by varying the parameter τ. This is how wavelet analysis provides a means for multi-resolution analysis. We can inspect the transformed signal at different scales by varying s.

We can now introduce continuous wavelet transform. We select a mother wavelet and compute correlation between the signal $x(t)$ and the basis functions:

$$w(s, \tau) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{s}} \Phi\left(\frac{t-\tau}{s}\right) \mathrm{d}t. \tag{1}$$

Wavelet coefficients w respond similarly to matching signal patterns as a tuning fork responds to matching sound—by gaining energy. Eq. (1) defined the continuous wavelet transform. For computational purposes, a discrete variant is introduced as

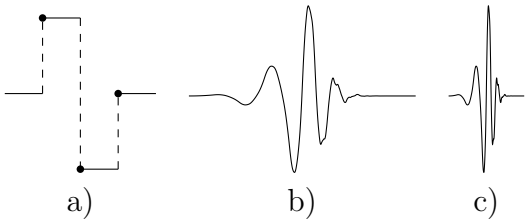

Figure 9: a) Haar mother wavelet, b) Daubechies-5 mother wavelet, c) scaled version of the Daubechies-5 wavelet.

$$w_{j,k} = \int_{-\infty}^{\infty} x(t)\Phi_{j,k}(t)\mathrm{d}t, \tag{2}$$

where according to the widely used dyadic sampling $s = 2^{-j}$, $\tau = k2^{-j}$, and $\Phi_{j,k} = s^{j/2}\Phi(2^j t - k)$. The variables j and k are integers.

So far we have discussed wavelets using graphs of mother wavelets and correlations with the signal under study. It can be shown that the discrete wavelet transform can also be formulated as a filtering problem. This implementation, the filter bank method, is computationally very efficient. Mallat was the one to introduce this algorithm [58].

One run through the filter bank breaks the signal $x[t]$ into two parts: approximation $x^a[t]$ and details $x^d[t]$. Approximation is the coarse scale that contains mainly the trend of the signal. More specifically, it contains frequencies smaller than half the bandwidth of the signal. Details capture the fine scales, or the upper half of the bandwidth. Hence, each run divides the bandwidth of the signal in two. At step $j + 1$, we have

$$x_j^a[t] = x_{j+1}^a[t] + x_{j+1}^d[t]. \tag{3}$$

The original signal can be restored from the approximations and details. Fig. 10 shows an example of wavelet coefficients at different levels. We also see how the high-frequency burst is captured, and remains well focused in time, by the detail part, while the approximation part provides the trend of the signal.

The major drawback of the basic discrete orthogonal wavelet transform is that it is not time-invariant, meaning that coefficients of a delayed signal are not a delayed version of the coefficients of the original signal. In signal processing, this would mean that performance of a detector depends on when the signal arrives at the detector. Obviously, such behaviour is detrimental to the credibility of the detector. The stationary variant of orthogonal wavelet transform was first introduced by Pesquet et al. [59]. All wavelet transforms in this thesis are carried out using the time-invariant method.

3.2.2 The Hilbert transform

Johansson has presented a thorough review of different aspects of the Hilbert transform [60]. The Hilbert transform $\hat{f}(t)$ of a function $f(t)$ is defined by

$$\hat{f}(t) = \frac{1}{\pi}\mathrm{P.V.}\int_{-\infty}^{\infty} \frac{f(\xi)}{t-\xi}\mathrm{d}\xi, \tag{4}$$

where P.V. denotes the Cauchy principal value of the integral. The transformation stated above arises in several fields of mathematics. In signal processing, it is most often used when creating an analytical signal from a real signal. Analytical form $z(t)$ of function $f(t)$ is given by

$$z(t) = f(t) + i\hat{f}(t). \tag{5}$$

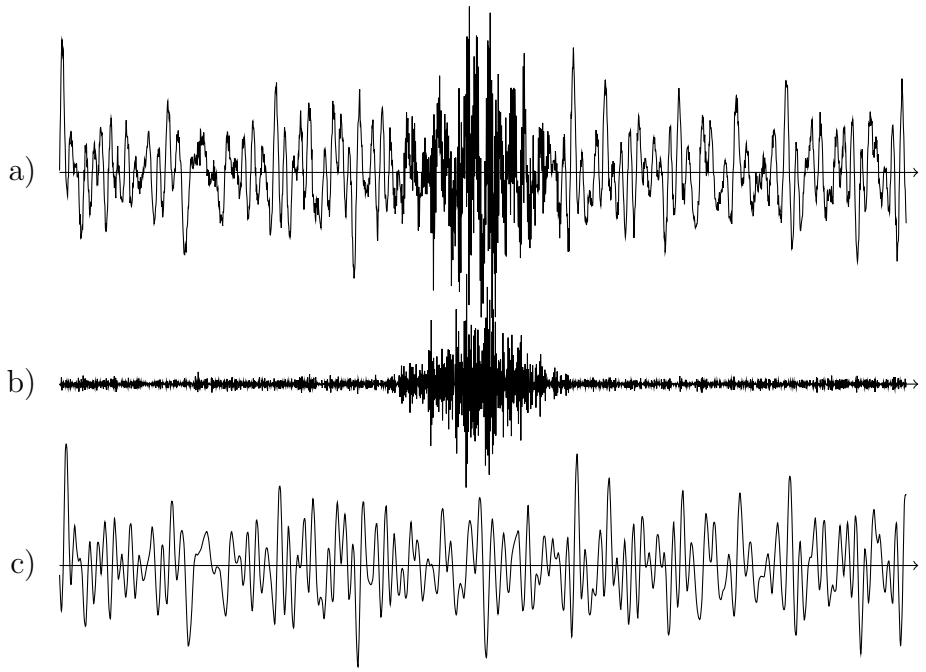

Figure 10: An illustration of stationary wavelet transform with Daubechies-5 mother wavelet. a) Signal containing a high-frequency burst superposed on a stationary signal with Gaussian noise. b) First detail level captures most of the noise and the burst. c) Second-level approximation. Units in this figure are arbitrary.

Analytical signals can be written using complex exponential function:

$$z(t) = A(t)e^{i\phi(t)}, \tag{6}$$

where

$$A(t) = \sqrt{f(t)^2 + \hat{f}(t)^2} \tag{7}$$

and

$$\varphi(t) = \arctan\left(\frac{\hat{f}(t)}{f(t)}\right). \tag{8}$$

Knowing the phase of the signal, we can compute the instantaneous angular frequency as

$$\omega(t) = \frac{\mathrm{d}\phi(t)}{\mathrm{d}t}. \tag{9}$$

Frequency in Hz corresponds to $\omega/2\pi$. In the case of multi-component signals, instantaneous frequency represents the local frequency averaged over a few samples.

Above we have discussed continuous functions. EEG, however, is a discrete signal. Deriving the discrete variant of the Hilbert transform is somewhat involved. Suffice it to say that the algorithm utilizes fast Fourier transform.

3.2.3 Signal power

Signal power is one of the simplest signal features. In this thesis, we compute signal power from wavelet transformed EEG signal. Average signal power over N samples is given by

$$P[t] = \frac{1}{N} \sum_{k=t-N+1}^{t} x[k]^2. \tag{10}$$

To obtain meaningful values for signal power, we must remove the DC component before computation. This is done by applying a high-pass filter before the wavelet decomposition.

In the analyses, we will use the instantaneous frequency as defined in the preceding subsection and base-10 logarithm of signal power.

3.3 Evaluation methods

Evaluating the performance of an algorithm is anything but a straightforward task. To objectively assess its performance, we must put it into numbers. However, the numbers we choose to present can have a drastic effect on the interpretation. Thus, special attention must be paid on which error measure and performance measure we report.

In general, there are two approaches to assessing how well an algorithm agrees with expert opinion. The first method is to think of detections as binary events, not paying attention to their durations. With this methodology, if the detection made by the algorithm overlaps with that made by the expert, we count it as a true positive detection (TP). Similarly, a detection that does not overlap with expert's annotation counts as a false positive detection (FP). This consideration gives us two performance measures, any-overlap sensitivity ($\mathrm{Se_{AO}}$) and false positive rate (FPR):

$$\mathrm{Se_{AO}} = \frac{\text{Number of TP}}{\text{Total amount of expert annotations}} \tag{11}$$

$$\mathrm{FPR} = \frac{\text{Number of FP}}{\text{Duration of the recording, annotated events excluded}} \tag{12}$$

The obvious shortcoming with these measures is that they do not measure how well the detections cover the expert annotations or how long the false positive detections are. In the extreme situation, an algorithm that marks the whole recording with intermittent annotated seizures as a single detection would yield $\mathrm{Se_{AO}} = 1$ and $\mathrm{FPR} = 0$, regardless of the number of annotations and their durations. For a seizure-free recording of duration t, such an algorithm would have $\mathrm{FPR} = 1/t$. Throughout this thesis, the unit of FPR is h^{-1}.

These shortcomings can be addressed by using integral measures. Instead of merely checking whether the detection and the annotation overlap, we compute how much they overlap. Similarly, for the false positive detections we compute how big a proportion of detections occurring outside expert annotations take up of the

seizure-free part of the recording. Now we can introduce integral sensitivity ($\mathrm{Se}_{\mathrm{Int}}$) and integral specificity ($\mathrm{Sp}_{\mathrm{Int}}$):

$$\mathrm{Se}_{\mathrm{Int}} = \frac{\text{Duration of TP}}{\text{Duration of expert annotations}} \tag{13}$$

$$\mathrm{Sp}_{\mathrm{Int}} = \frac{\text{Duration of FP}}{\text{Duration of the recording excluding the annotated parts}} \tag{14}$$

With the integral measures, an algorithm that marks the whole recording as a single detection would perform ideally in terms of $\mathrm{Se}_{\mathrm{Int}}$ but the value of $\mathrm{Sp}_{\mathrm{Int}}$ would reveal the culprit. These measures are informative if the algorithm is designed to detect the whole annotated epoch. But if the algorithm is designed to detect only the on-set or the off-set of the event, these performance measures are misleading.

Wilson has presented a discussion with some novel performance measures [61]. In publications, the most common numbers to present are $\mathrm{Se}_{\mathrm{AO}}$ and FPR.

What complicates the performance evaluation even more than the selection between different measures is the lack of rock-solid ground truth. When annotating seizures, there is a considerable inter-expert discrepancy [35]. The differences are most pronounced when annotating on-sets and off-sets of seizures. In author's opinion, the point of the most active ictal activity can be rather easily spotted, but annotating the gradual on-set is rather ambiguous. This is extremely problematic when considering $\mathrm{Se}_{\mathrm{AO}}$ as a performance measure. A detection occurring just before or after the expert's annotation might, actually, be a correct detection, but does not count as a true positive detection. However, we cannot justify changing the on-set and the off-set annotations, because such a detection may be caused by some other change in the EEG than seizure on-set or off-set.

We are attempting to develop a detector that is sensitive for evolution in certain signal characteristics. Evolution occurs typically throughout the seizure, but when isolated seizures start to merge and the patient proceeds towards SE, the amount of evolution diminishes [62]. If we consider $\mathrm{Se}_{\mathrm{Int}}$ as performance measure, we should expect to see poor performance in prolonged ictal activity.

To address the challenges discussed above, we present one more performance measure. From the clinical standpoint, the most important event is the first seizure. If we can detect this event within a certain time frame, the staff can intervene and further complications can be avoided. Since it is not critical that the method detects the exact on-set of the annotated event, we relax the limits a bit by introducing first event sensitivity ($\mathrm{Se}_{\mathrm{AO}}$ 1st). Let Δt be the interval that spans from 3 min before the first on-set annotation to 10 min after the on-set annotation. For each patient, define:

$$\mathrm{Se}_{\mathrm{AO}}\ 1\mathrm{st} = \begin{cases} 1, \text{ if there is a detection during } \Delta t \\ 0, \text{ otherwise} \end{cases} \tag{15}$$

Differences between the performance measures introduced here are summarized in Fig. 11. In this example depicting a 1 h recording, there are two events annotated by the expert. Three exemplary algorithm outputs are presented and the extend of

Δt is illustrated. Table 8 lists the performance indicators' values for the example algorithms. By merely looking at the performance values, it is hard to say which algorithm performs the best.

All these issues make the performance evaluation a difficult task. On the one hand, putting performance in numbers is the only way to objectively assess how well the algorithm works. On the other hand, the performance numbers can oversimplify and mislead the audience and the researchers. We can conclude that since no performance measure alone is ideal for our purposes, we will have to, at least to some extent, trust our subjective assessments when comparing different methods. Extra care should be taken when considering $\mathrm{Se_{AO}}$ 1st, since the detection might actually arise from an artefact.

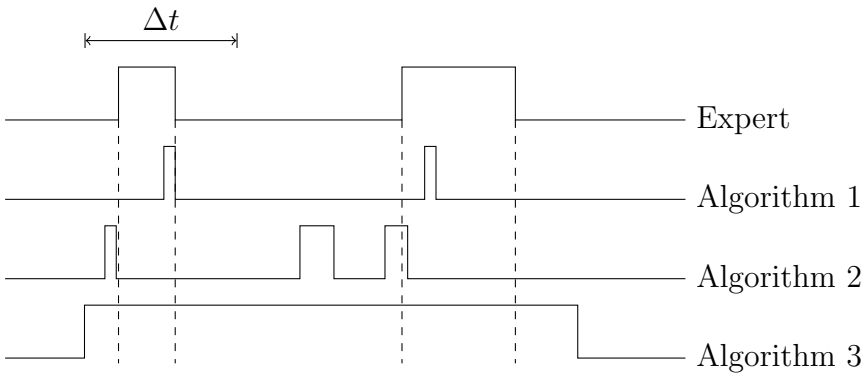

Figure 11: Examples of different algorithm outputs and performance measures.

Table 8: Performance of example algorithms presented in Fig. 11

	$\mathbf{Se_{AO}}$	**FPR**	$\mathbf{Se_{Int}}$	$\mathbf{Sp_{Int}}$	$\mathbf{Se_{AO}}$ **1st**
Algorithm 1	1	0	0.13	1	1
Algorithm 2	0.5	2	0.03	0.88	1
Algorithm 3	1	0	1	0.36	1

4 EEG feature space

4.1 EEG preprocessing

The preprocessing steps are summarized in Fig. 12. We start with a single-channel signal. The signal is first down-sampled to 128 Hz. Filtering is performed using a low-pass equiripple FIR filter with stop-band at 50 Hz and a high-pass equiripple FIR filter with stop-band at 0.25 Hz. Filters were designed in MATLAB and had lengths of 94 and 880 samples, respectively. Stop-band attenuation was set to 70 dB. Filter lags are accounted for in the processing. Filtering effectively removed the mains interference and the DC component.

In the second phase, artefact exclusion criteria are applied. When rejecting artefacts, we prefer too strict criteria to too loose. Artefact detection is done to ensure that what the algorithms process is likely to be related to the patient's brain activity. An amplitude criterion is used for detecting mechanical artefacts. In addition, loose electrodes are detected by monitoring signal power in the 32–64 Hz band.

The artefact-free signal is then transformed using stationary wavelet transform. For further analysis, we select wavelet approximations from 0–16 Hz band. Using this representation, we compute the features IF and $LOGPOW$. The features are estimated in 1 s non-overlapping windows. Having filtered the features using a median filter of length 21, we arrive at the final features.

4.2 Noise susceptibility of computed features

Fig. 13 shows the relationship between noisy signal amplitudes and computed $LOGPOW$ values. We use the mathematical notion of signal amplitude, i.e., half of the peak-to-peak difference. Simulated values were obtained by using 3.5 Hz sinusoids with added Gaussian noise such that SNR $\approx$ 6 dB. In the figure, SNR is kept constant and the amplitude of the carrier signal is varied. The feature performs well even in the presence of small noise.

Peak-to-peak EEG difference below 10 µV is considered suppressed [11]. Such epochs will not yield any further analysis. From Fig. 14 we can see that this limit maps to $LOGPOW \approx 1.7$. Contrary to Fig. 13, Fig. 14 presents the data without added noise.

While the signal power feature is robust against noise, the frequency feature is not. Simulation results are shown in Fig. 15. The simulations were carried out with Gaussian noise ($\mu = 0, \sigma^2 = 3\,(\mu\mathrm{V})^2$). In this figure, the absolute amount of noise was kept constant, meaning that the SNR varies with the carrier signal amplitude. The noise parameters were selected according to a report describing the amount EEG electrode noise [63]. The selected parameters represent typical values for good electrical contact. Using this noise level, the IF readings become unreliable when $LOGPOW < 2$.

Features have been computed using a 1 s window. The features are, to a large extent, performing well with different frequencies as long as there are several cycles

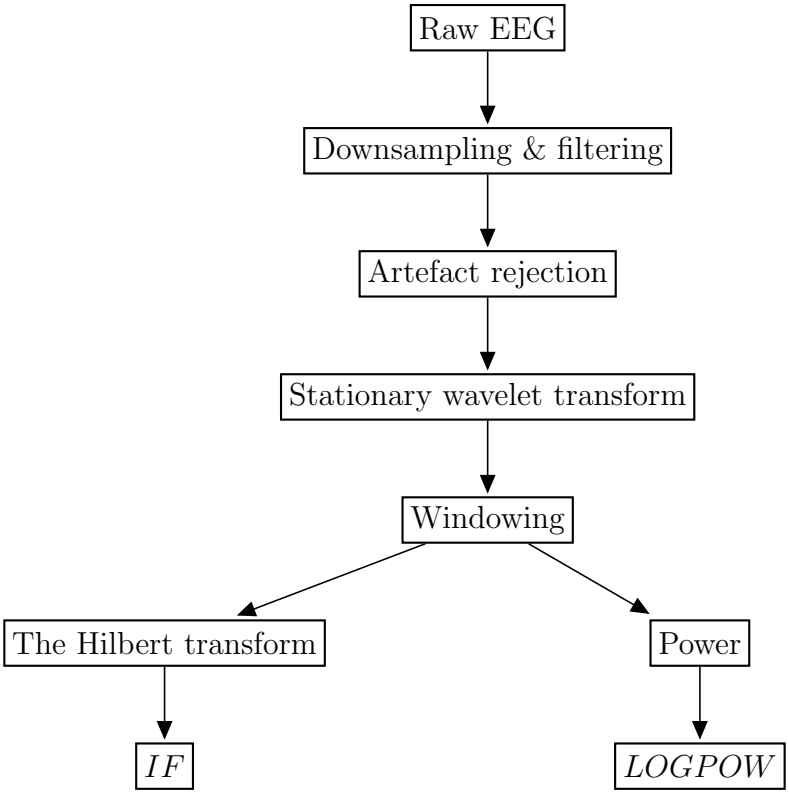

Figure 12: Preprocessing steps.

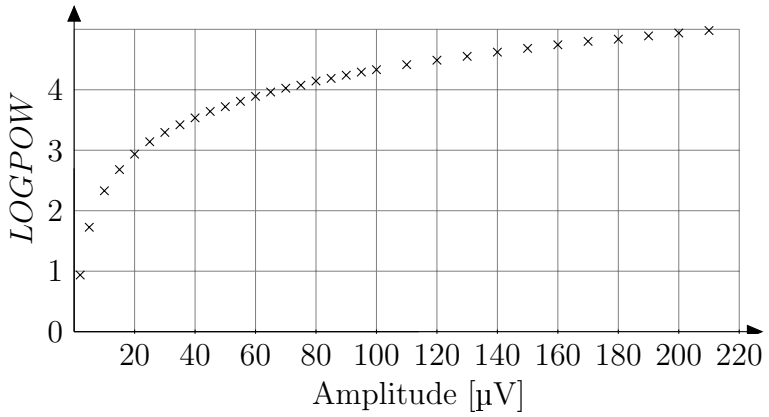

Figure 13: Relationship between signal amplitude and *LOGPOW* values. Values obtained from 3.5 Hz sinusoid with added Gaussian noise.

within the computation window. But when using a 1 s window, frequencies lower than 1 Hz will not yield correct *LOGPOW* values. This effect is seen in the 0.5 Hz signal in Fig. 15. The computed *LOGPOW* values are translated a bit leftwards.

However, the noise susceptibility of IF is not a function of the computation window length. Increasing the length of the computation window would give more

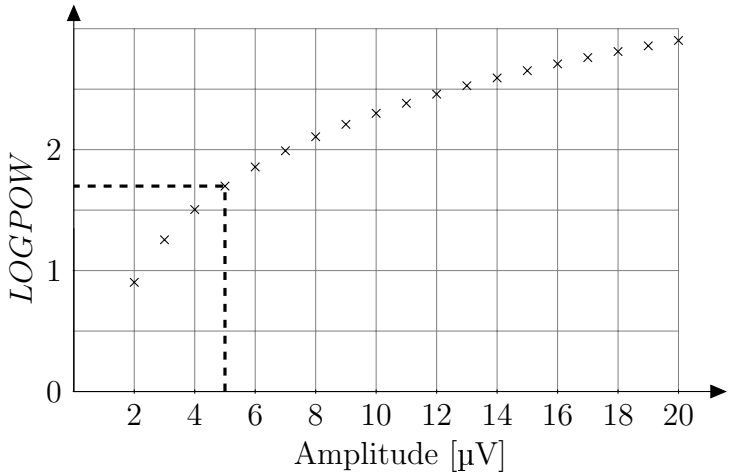

Figure 14: Amplitudes of noise-free sinusoids and the corresponding $LOGPOW$ values. Clinically considered limit for suppression yields $LOGPOW \approx 1.7$.

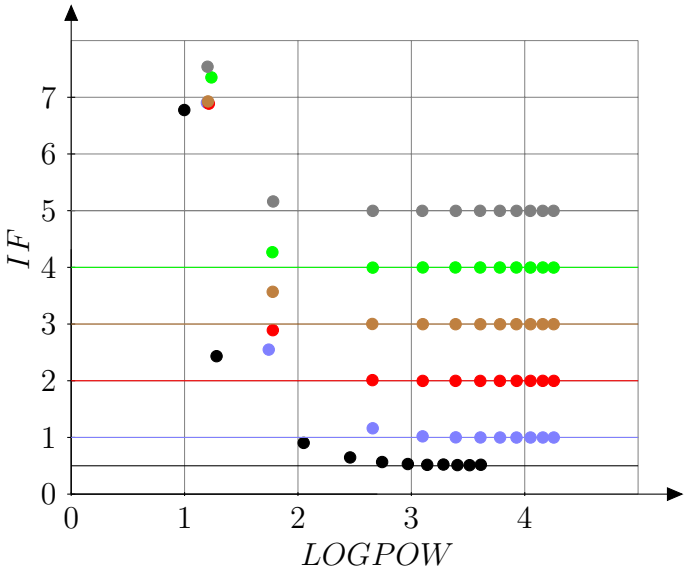

Figure 15: Performance of IF variable in the presence of Gaussian noise. Straight, coloured lines show the true frequency of the signal while circles depict the computed IF values at different power levels. In this figure, the amplitude of the sinusoid is varied and noise is kept constant. Distortions appear when $LOGPOW \leq 2$ and they are most pronounced in low frequencies.

reliable results in low frequencies in terms of $LOGPOW$ but it would also degrade temporal resolution.

The discussion above leads to the conclusion that as signal power diminishes below a certain limit the frequency variable's readings are not to be trusted. This is a crucial piece of information when investigating different algorithms and sources

of their possible shortcomings. The clinical limit of EEG suppression and the lower bound of signal power that yields reliable frequency estimates are, luckily, almost the same. When designing the algorithms, we have selected to exclude epochs where the feature values might not be reliable and EEG is close to the limit of suppression. At this stage the exclusion is based on binary logics. If EEG is close to the limit of suppression, this causes unwanted jumps in feature values. A more sophisticated way to handle EEG that is close to the suppression limit would be to design a spline that would make the transition smooth.

Low-frequency signals are more susceptible to noise than high-frequency signals. This is mostly due to the windowing parameters we selected. It is a widely appreciated fact that low-frequency signals need more samples for reliable analysis than high-frequency signals. One way to treat different frequencies equally would be to consider the amount of cycles in the computation window.

4.3 Prototype seizure

Fig. 16 shows an example of EEG feature traces during an evolution seizure. This single-channel example highlights the distinctive phases of a seizure and how the changes are reflected in feature values. Before seizure on-set, the background activity is stable. During seizure, we see first a clear increase in frequency and in power, followed by a further increase in power and slowing of EEG. Finally, as the seizure wanes, we see post-ictal suppression. See Fig. 7 for the EEG of this seizure.

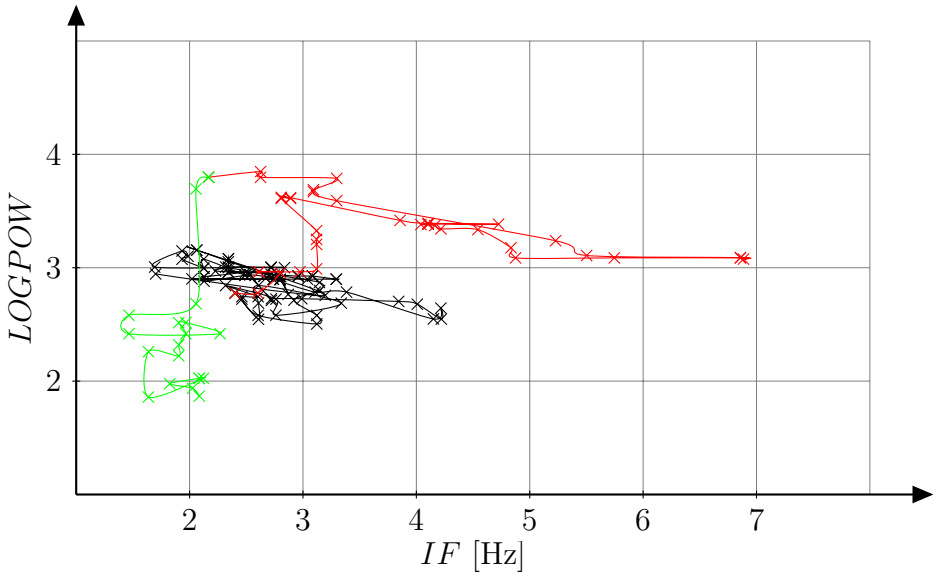

Figure 16: A prototype seizure. The black points represent feature values before seizure, the red points correspond to the time marked as seizure by a clinician, and the green points refer to the post-ictal period. Data points are sampled at 1 Hz. This is an actual seizure recorded on T_4–C_z. See Fig. 7 for the EEG tracing of this seizure.

We will use this prototype seizure to show how different algorithms approach the problem of detecting seizure-related evolution. It should be noted that while this example seems like an easy one to be detected, seizures in reality come in all shapes and sizes. This example is used merely for illustration purposes and by no means represents a template. Finally, it should be noted that even though we use single-channel data for visualization here, the algorithms process multichannel data, allowing for comparisons between brain regions.

It is insightful to pay attention to how the specialist has annotated the seizure. Feature values at the on-set cannot be distinguished from the background, but the expert has already seen ictal patterns. In striking difference with automated detection systems, specialists often review data rewinding forward and backward. This allows them to spot the most easily detectable phase of seizure and then to rewind back to fine-comb the channels to find the first indications. Same applies for the end mark. Automated systems, however, should make annotations in real time, without the possibility of returning back in time. For this reason, trying to optimize the system to follow specialist's annotations too strictly may not be a reasonable goal for the development project.

5 Developed algorithms

This section describes the algorithms that were developed in the course of this thesis. Three algorithms were developed using a slightly different rationale in each case. The three proposed algorithms use the same EEG features.

5.1 Algorithm I: Path length

The first attempt at detecting evolution arises from the rationale that as the seizure proceeds, the steps in the feature space do not cancel out as much as they do during inter-ictal epochs or in non-epileptiform EEG. Fig. 17 illustrates the idea. Since consecutive steps do not cancel out, with proper filtering and computation parameters, we should be able to derive an indicator that describes the length of the feature trace as perceived by the human eye.

It is clear from the example seizure that if we compute normalized path length in a given window, there should be an increase in the variable after seizure onset followed by decrease as the last seizure-related data points slide outside the computation window.

Let $x[t]$ and $y[t]$ be the features that are being traced and $\vec{F}[t]$ be a vector whose components are $x[t]$ and $y[t]$. Define difference over m samples as

$$\Delta\vec{F}_m[t] = \vec{F}[t] - \vec{F}[t-m]. \tag{16}$$

For the computation, we need to define the computation window length N. Define normalized path length (PL) as

$$PL[t] = \frac{1}{N}\sum_{n=t-N+m+1}^{n=t} ||\Delta F_m[n]||_2 \,. \tag{17}$$

Both m and N were at first tuned by hand. The parameter m means the number of samples over which the difference is sought. For example, using $m = 5$ means that differences are considered over five samples and we obtain five difference values for each 5 s block. After the computation, we apply a median filter of length $m + 1$. Increasing m decreases noise in PL.

The value of m was tuned by hand to $m = 5$. Computation window length is an important parameter that was left to be optimized in Sect. 6.

To complete the path length algorithm, we compute a weight vector $w[t]$ whose purpose is to make the seizure-related evolutions more pronounced in the feature values while attenuating effects of evolutions that are not likely to be seizure-related. Fig. 18 illustrates the weight computation. Let

$$\Delta\vec{f}_m[t] = \frac{\Delta\vec{F}_m[t]}{\left|\left|\Delta\vec{F}_m[t]\right|\right|_2}. \tag{18}$$

The weighting term is then obtained as

$$w[t] = \max\left(0, \Delta\vec{f}_m[t]\cdot\vec{p}\right), \tag{19}$$

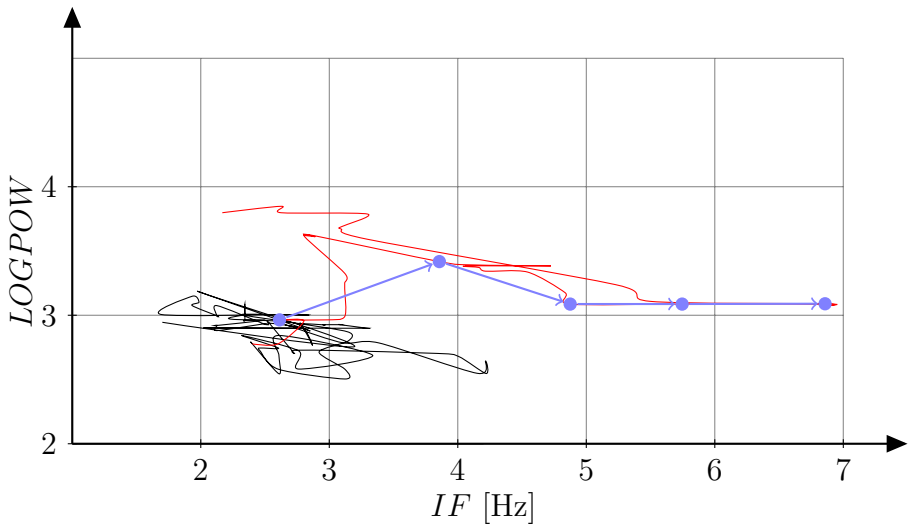

Figure 17: Illustration of path length computation. During the seizure the steps do not cancel out as much as during normal activity. In the ictal phase they are targeted towards higher frequency and amplitude values. Arrows show some vectors $\Delta\vec{F}_m$ for different t.

where $\vec{p}$ represents the direction associated with seizure related evolutions. Translating the clinicians' criteria into mathematics, we set $\vec{p}$ at an angle of $\pi/4$ with the positive x-axis. We set $||\vec{p}||_2 = 1$, yielding weights in $[0, 1]$. Final, weighted path length values are obtained as

$$PLW[t] = w[t] * PL[t]. \tag{20}$$

Fig. 19 shows the tracing of PLW during the prototype seizure. It is interesting to note that the values start rising already before the annotated start of the seizure. PLW is suitable for detecting seizures with evolution right from the beginning of the measurement because it does not need history data.

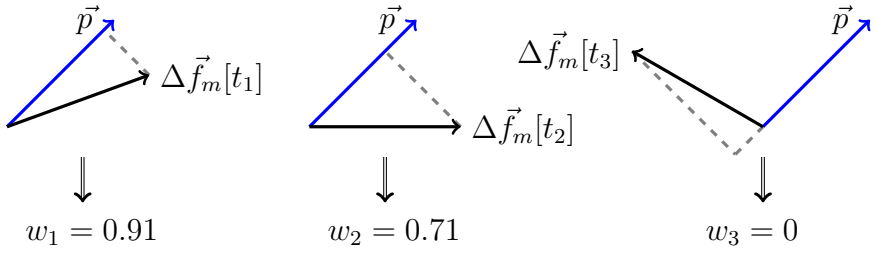

Figure 18: Illustration of how weighting factors for path length computation are obtained. Gray dashed lines mark the projections of normalized differences over m samples (black) to the vector $\vec{p}$.

To complete the detector, we need to define a threshold for the indicator. This is done for PLW as well as for the other algorithms in Sect. 6.

5.2 Algorithm II: Random walk

The second algorithm is based on the idea that when there is no seizure activity, the feature tracing resembles the path of a random walker. Each step seems to be randomly drawn from an autoregressive process At seizure on-set, the process changes from random to organized. Steps are no longer drawn from a stochastic constant-mean process, but seem to correlate in direction and size. This algorithm attempts at quantifying the change.

We start by defining one or more histories N. These histories form the memory of the process. Using the histories, we define the average step as

$$\vec{a}[t] = \frac{\sum_{n=t-N}^{t} \left|\Delta \vec{F}_m[n]\right|}{N} \tag{21}$$

For each step, we compute an index indicative of the likelihood of the step being ictal. Steps are computed considering the difference between the latest point and the average value within a block of size b that is N seconds in the past. Averaging is used for making the algorithm more robust. A step is computed by

$$\bar{F}_N[t] = \frac{1}{b} \sum_{k=1}^{b} \Delta \vec{F}_{N-1}[t]. \tag{22}$$

An index RWS representing the amount of steps taken by the random walker can then be defined as

$$RWS = \left|\left| \left(\frac{\bar{F}_{N,1}[t]}{a_1}, \frac{\bar{F}_{N,2}[t]}{a_2} \right) \right|\right|_2 \tag{23}$$

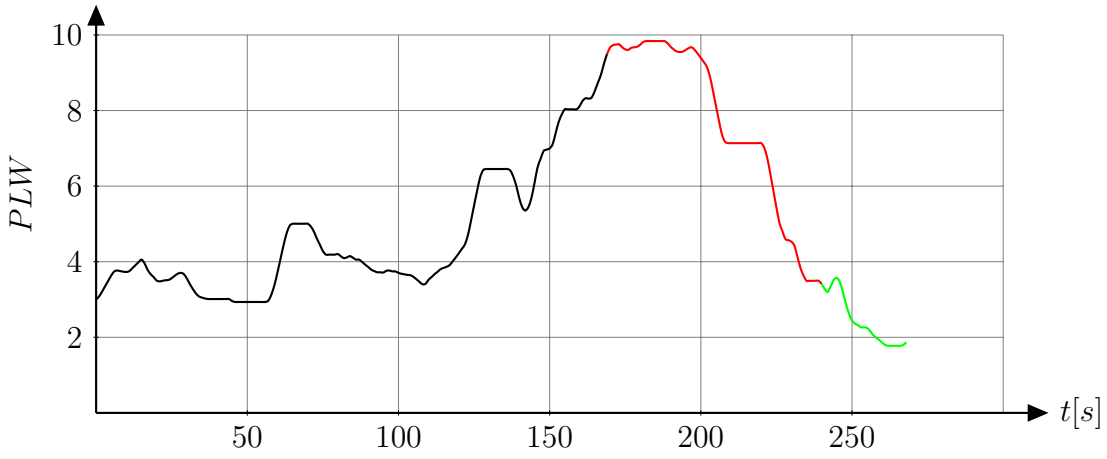

Figure 19: Output of the algorithm I, PLW, showing the variable's behaviour before seizure (black), during seizure (red), and during post-ictal suppression (green).

where subscripts 1 and 2 refer to the components of the vector.

Simply put, we compute how many average steps it would take in both x and y directions to represent the difference between the last step and the average value N seconds in the past. If the number is very small, we can conclude that the current step does not stand out from the history data. A large value means that we can immediately recognize the step as not being a part of the background activity. While one could present arguments for using the norm $||\cdot||_\infty$, we have used the Euclidean norm.

The original idea behind this algorithm was to consider the tracings of the features as a path of a random walker. A random walker takes steps whose length and direction are randomly selected from a given distribution. We gathered background EEG and composed what was a best guess of the step distribution. Then, given a step, one could easily check how probable would it be for the random walker to take a step that would be longer than the one actually taken. This probability tends to zero as step sizes grow. With small steps, the probability of a longer step is large, an indication of the fact that the step does not stand out from the background.

The original idea was found to be too noisy and we let go of the probability distribution and focused on the number of average steps it would take to make the current step. This algorithm, being also computationally favourable, proved to be more robust than the original one. RWS provides a similar trending feature as PLW. Fig. 21 shows how the indicator behaves during the prototype seizure.

Adjustable parameters are limited to the length of the history N and the scope of the step size m. The latter was selected by similar arguments as with the previous algorithm, and N was left for later optimization. Also, we can choose to use a fixed

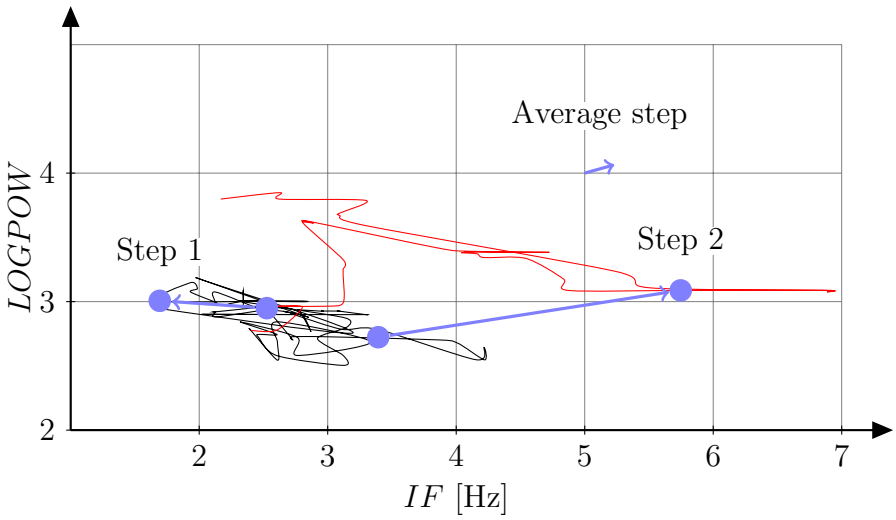

Figure 20: Illustration of algorithm II. The average step has been defined from the background activity. Two exemplary steps depict the idea of the algorithm (see also Fig. 21).

step size instead of adaptive steps when performing the comparison. Fixed step size requires the determination of one more parameter, the representative step of the background activity.

5.3 Algorithm III: Convex hull

Like the random walk algorithm, the convex hull algorithm is designed to detect changes relative to background. The algorithm keeps track of history data with detected evolutions excluded. Using these data, a borderline is constructed. The key idea is that as long as the feature vectors stay within the boundaries, the changes do not count as seizure-related changes. We have used a background window of maximum 3 min.

When a seizure starts, the computed EEG features, desirably, also change. Depending on the seizure and on the features used, the changes can be minuscule or very large. The main point is that they do exhibit a change compared to the background. Defining background activity and its boundaries is a demanding task as such. We use the concept of convex hull to establish the boundaries of the background activity. Fig. 22 shows a simplified example of how the algorithm works.

Let $X = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k | \vec{x}_{1\ldots k} \in \mathbb{R}^n, k \in \mathbb{N}\}$. One way of defining the convex hull of set X is

$$\mathrm{conv}(X) = \left\{ \sum_{i=1}^{k} a_i \vec{x}_i | \vec{x}_i \in X, a_i \in \mathbb{R}^+, \sum_{i=1}^{k} a_i = 1, i \in \mathbb{N} \right\}. \tag{24}$$

The notion of convex hull can be easily understood if one considers the points of X scattered in a two-dimensional space and a rubber band that is stretched so that whole X is inside the rubber band. When the rubber band is released, it takes the form of $\mathrm{conv}(X)$. The same concept is readily generalized to larger dimensions. Convex analysis is a well-established branch of mathematics and the problem of

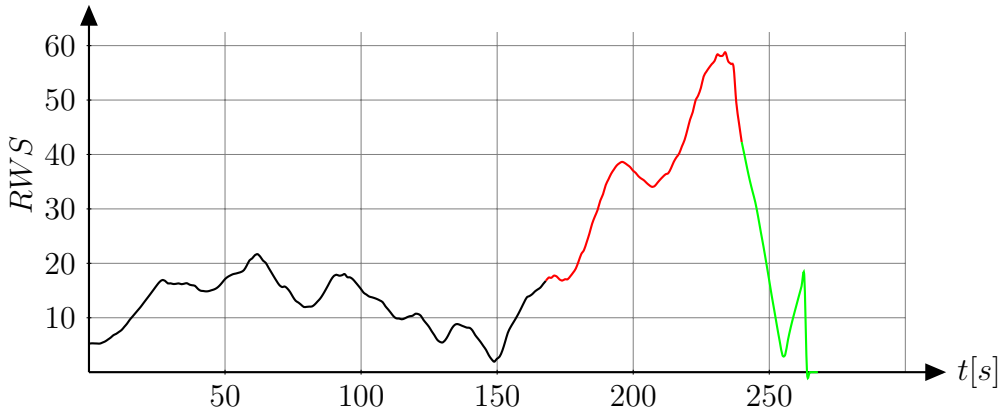

Figure 21: Output of the algorithm II, RWS, during the prototype seizure. The computation was carried out with adaptive step sizes and a 180 s history.

constructing the convex hull of a given set is studied thoroughly. In our application, we use the *quickhull* algorithm [64].

Each time the feature vectors breach the hull, an elementary detection is made. This elementary detection is subject to further tests to check whether it is seizure-related or not. First, we compute the angle ν between the center-of-mass of the background data and the most recent data point. Only a certain range of angles is acceptable for seizure-related evolution. This is a way of translating the clinicians' criteria for seizures into a set of rules.

The second test considers the duration of the elementary evolution. There is a time threshold $t_{\min}$ that the feature vectors must remain outside the hull. If the elementary detection is very brief, it will be discarded and considered normal background activity. Discarded detections, as they now are a part of the background activity, are used for updating the hull.

The final criterion before a detection is made is the maximum distance d to the hull reached during the elementary detection. We use a threshold $d_{\min}$ to exclude elementary detections that have remained very close to the hull.

Detection can end when feature vectors return inside the hull, close enough to its border, or when maximum detection duration $t_{\max}$ is exceeded. When the detection ends, the hull is reset.

The output of the algorithm is a vector CHS showing how many consecutive steps have been taken outside the convex hull. This is in stark contrast with the previously introduced algorithms that produce a trend. Parameters $t_{\min}$ and $d_{\min}$ will be optimized like with other algorithms while the range of ν is selected based on commonly accepted seizure criteria. An example output of the algorithm is shown in Fig. 23. The decision-making structure of the algorithm is summarized in Fig. 24. This algorithm is clearly more complex than the previous two algorithms.

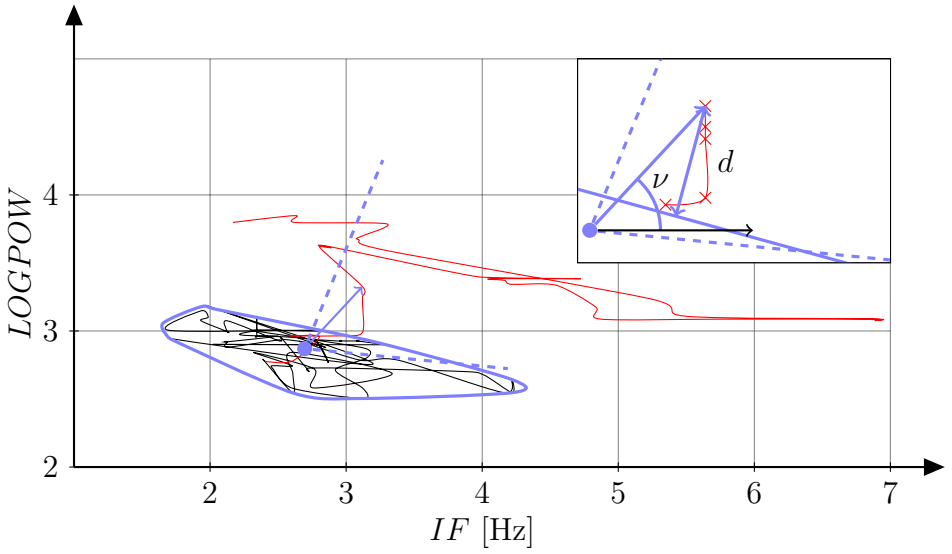

Figure 22: The background activity is confined in the convex hull (blue). Soon after seizure on-set, the feature vectors leave the convex region. Dashed lines show the range of directions relative to the center-of-mass that are considered acceptable for seizure-related evolution. In the close-up figure, distance d is evaluated after five consecutive steps outside the convex hull ($t_{\min} = 5$).

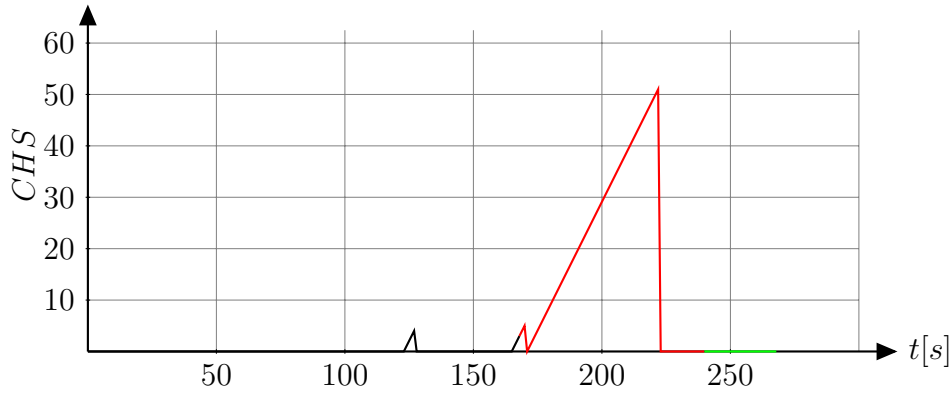

Figure 23: An example of the output of the algorithm III, CHS, during the prototype seizure.

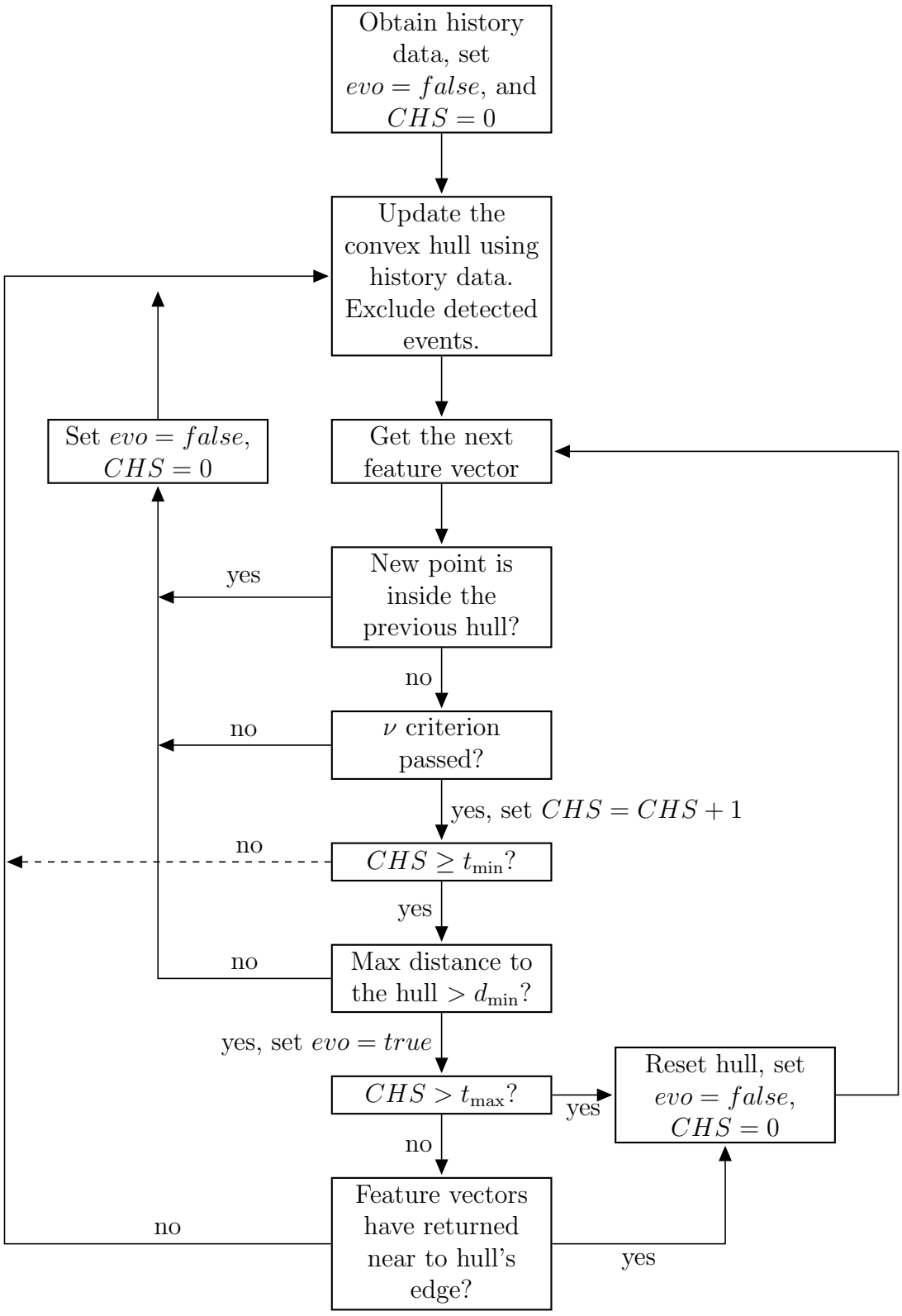

Figure 24: Flow chart of algorithm III. The binary variable *evo* is true when seizure activity is detected and false otherwise.

6 Results

Algorithms I and II produce trends. High indicator values mean that the effect measured by the method is pronounced in the data. For example, path length has a baseline value that corresponds to the amount of normal evolution in the EEG. Slightly increased amount of evolution in EEG yields a slightly higher path length value, and a large amount of evolution produces large output values. The random walk algorithm works in a similar fashion. For these algorithms, we have to define a threshold that allows us to separate ictal epochs from other data.

Algorithm III does not produce a trend. It uses more logics and tests than the other algorithms. As a consequence, this method cannot be analysed in the same way as the other algorithms. The output variable, CHS, depicts the number of steps outside the convex hull. After $t_{\min}$ steps additional tests are performed. Based on them, the counter is either reset or allowed to increase further. It follows that CHS value of $t_{\min}+1$ should count as a detection. Contrary to the other algorithms, higher values of CHS do not mean that it is more likely that there is an on-going seizure. It only means that tests have been passed and the hull has not yet been reset. For this reason we cannot apply similar thresholding techniques as we will be using for the other two algorithms.

In the first subsection we introduce the performance evaluation protocol. The rest of the subsections each address one algorithm. After optimizing the detection parameters, we present the results. At the end of each subsection we report problematic recordings. After presenting results for individual algorithms, we address the question of selecting the most prominent one of them by comparing performance values at both seizure detection and at avoiding false positive detections.

6.1 Performance evaluation protocol

All three proposed algorithms have internal parameters that should be optimized. That is, however, computationally not possible. Computations were carried out using a modern workstation with two quad-core processors running at 2.66 GHz and 3 GB of RAM. Using the whole database (see Table 6), 8 EEG derivations, and computing feature values at 1 Hz yields a running time of about 30 h for algorithm III. Algorithms I and II run in about 12 h and 15 h, respectively. Some time savings could be achieved by optimizing computation routines, but the fact remains that sweeping the entire parameter space of even one algorithm is a demanding task. All computations were carried out in MATLAB version 6.5 [65]. The available version of MATLAB could not employ the whole potential of modern multi-core technology.

Because we cannot use statistical methods to pick up the optimal parameters, we must resort to a manual selection procedure. In the following subsections we will present three variants for each algorithm. First version, (v_1), is selected so that specificity is prioritized. Using this configuration, we will learn which patients are the easiest ones to detect. We will also find out which of the non-seizure patients have detections despite the strict criteria. Such patients have to be addressed separately. Once the source of false detections is identified, we can complement the algorithm

with a routine that detects and excludes such epochs.

Second configuration (v_2) is designed so that sensitivity of the method is maximized, while keeping the amount of false detections at a reasonable level. With this set-up, we can learn which patients are likely remain out of the scope of the method. Again, these patients must be addressed separately, e.g., by spike detector.

Third configuration (v_3) is selected between the two extrema. This configuration represents a compromise and gives an idea of the realistic potential of the method at the current development stage. This version is not presented for algorithm III because of the method's computational burden.

There is a considerable amount of subjective judging when picking the representatives of each method. There is no reasonable way to pick the configurations automatically. The author has used the same criteria with all methods when selecting parameter values, but this remains a source of subjectiveness in the study.

We present several performance measures for each algorithm. False positive detections are reported only for patients with no annotated seizures. FPR_w refers to the duration-weighted average FPR. FPR_{pa} refers to non-weighted patient averages, obtained after pooling all recordings of each patient together. Se_{AO} 1st refers to Se_{AO} when considering relaxed first annotated seizure of the patient (see Fig. 11). Patient sensitivity is 1 for seizure patients that have any detections in Se_{AO} sense. In results, we present the value averaged over all seizure patients (Se_{AO} pt). We also report Se_{AO} when all seizures in the database are considered (Se_{AO} all szs), and the average of patients' Se_{AO} values (Se_{AO} pt szs). Integral specificities (Sp_{Int}) are reported in the same fashion as FPR values. Finally, we report integral sensitivity computed over all seizure records (Se_{Int}).

We have to investigate the performance values as whole. Concentrating too much on a single value can be misleading. The lack of one accepted performance value can, unfortunately, introduce a bit of vagueness in the arguments presented below.

6.2 Performance of algorithm I

Since we are not interested in short transients that cause the indicator to rise above the threshold, we need to establish a threshold also on the time axis. Based on clinical guidelines, we set the threshold to 10 s. Thus, for a detection to be counted, the indicator value shall exceed the threshold continuously for more than 10 s.

To obtain the three representative versions, we try out several threshold values. We start with v_2 using a low threshold. Next, we start gradually rising the threshold. We are looking for a reasonable compromise between FPR and Se_{AO}. The threshold for v_2 is set at the value from which further increase cannot be achieved without significantly deteriorating detection capability.

For v_1, we continue rising the threshold value until we find a limit that gives detections only on the most pronounced seizures. Threshold for v_3 is selected between the extrema.

Results obtained using this scheme are summarized in Tables 9 and 10. PL30 refers to a 30 s history, PL90 to a 90 s history, PL180 to a 180 s history, and PLAVE to the average of the three. The data show that PLAVE and PL30 are the most

promising variants of this algorithm. PLAVE v_3 achieves one of the highest Se_{AO} 1st values while still having overall FPR less than one. PL180 stands out as having the worst performance values. Integral specificities are close to one with all window lengths, implying that most false positive detections are short. In fact, because Int_{Se} values are also low, the true positive detections are in general much shorter than the annotated seizures.

Using the v_1 and v_2 set-ups, we can identify problematic patients. Seizure patients are considered problematic if their Se_{AO} or Se_{AO} 1st values in the sensitive set-ups are close to zero. From the group of non-seizure patients, we list patients with FPR > 5 or $Sp_{Int} < 0.9$. Problematic patients are listed in Table 11.

6.3 Performance of algorithm II

Algorithm II can be evaluated using the same procedure as the previous algorithm. RWL refers to 180 s adaptive window, RWSH to 90 s adaptive window, and RWLF

Table 9: Detection results for algorithm I. This table presents Se_{AO} values. V_1 refers to the specific set-up, v_2 to the sensitive set-up, and v_3 to the compromise set-up.

	PL180			**PL90**			**PL30**			**PLAVE**		
Ident.	$\mathbf{v_1}$	$\mathbf{v_2}$	$\mathbf{v_3}$	$\mathbf{v_1}$	$\mathbf{v_2}$	$\mathbf{v_3}$	$\mathbf{v_1}$	$\mathbf{v_2}$	$\mathbf{v_3}$	$\mathbf{v_1}$	$\mathbf{v_2}$	$\mathbf{v_3}$
case016	0	0.33	0.33	0	0.67	0	0.33	0.67	0.33	0	0.33	0.33
case035	0	0	0	0	0	0	0	0.08	0	0	0	0
case098	0	0.33	0	0	0.33	0	0	0.33	0	0	0.33	0
case101	0	0.24	0	0	0.32	0.04	0	0.24	0.04	0	0.16	0
case102	0	0	0	0	0	0	0	0	0	0	0	0
case103	0	0	0	0	0	0	0	0.20	0	0	0.10	0
case104	0	0	0	0	0	0	0	0	0	0	0	0
case106	0	0	0	0	0	0	0	1	0	0	1	0
case109	0	1	0	1	1	1	0	1	1	0	1	0
case119	1	1	1	1	1	1	0	1	0	0	1	0
case122	0	0.67	0	0.33	1	1	1	1	1	1	1	1
case126	0	0.90	0.10	0.70	0.90	0.90	0.90	1	0.90	0.90	0.90	0.90
case130	0	0	0	0	0	0	0	0	0	0	0	0
case144	0.10	0.90	0.10	0.70	0.90	0.90	1	1	1	0.80	1	1
case156	0	0	0	0	0	0	0	0	0	0	0	0
case157	0.03	0.03	0.03	0	0.03	0	0	0.13	0	0	0.03	0
case159	0	0	0	0	0	0	0	0	0	0	0	0
case177	0.17	0.42	0.33	0.33	0.42	0.33	0.25	0.42	0.33	0.25	0.42	0.33
case178	0	0	0	0	0	0	0	0.20	0	0	0	0
Thresholds:												
	24.2	17	22.4	30	20.5	27	55	32	30	36	24	50

Table 10: Statistics for algorithm I.

		$\mathbf{FPR_w}$	$\mathbf{FPR_{pa}}$	$\mathbf{Se_{AO}}$ 1st	$\mathbf{Se_{AO}}$ pt	$\mathbf{Se_{AO}}$ all szs	$\mathbf{Se_{AO}}$ pt szs	$\mathbf{Sp_{Int,pa}}$	$\mathbf{Sp_{Int,w}}$	$\mathbf{Se_{Int}}$
PL180	v_1	0.43	0.61	0.05	0.21	0.04	0.07	0.99	0.99	0.01
	v_2	2.01	3.07	0.45	0.53	0.27	0.31	0.95	0.95	0.21
	v_3	0.66	1.04	0.20	0.32	0.07	0.10	0.99	0.99	0.03
PL90	v_1	0.43	0.68	0.25	0.32	0.16	0.21	0.99	0.99	0.04
	v_2	2.42	3.64	0.40	0.53	0.29	0.35	0.96	0.96	0.18
	v_3	0.74	1.19	0.30	0.37	0.21	0.27	0.99	0.99	0.09
PL30	v_1	0.38	0.78	0.30	0.26	0.19	0.18	1.00	1.00	0.05
	v_2	2.78	4.36	0.60	0.74	0.35	0.43	0.98	0.98	0.16
	v_3	0.56	1.04	0.35	0.37	0.22	0.24	1.00	0.99	0.07
PLAVE	v_1	0.36	0.97	0.25	0.21	0.17	0.16	0.99	0.99	0.04
	v_2	2.07	3.52	0.55	0.63	0.28	0.38	0.98	0.98	0.13
	v_3	0.84	1.60	0.50	0.26	0.20	0.19	0.99	0.99	0.07

Table 11: Detection problems with algorithm I.

Group	**Identifiers**
Seizure patients	case035, case102, case103, case104, case130, case156, case157, case159, case178
Non-seizure patients	case006, case015, case017, case044, case045, case050, case056, case066, case090, case096, case116, case140, case142, case146, case180

to 180 s fixed window.

Thresholds are set using a similar method as for the previous algorithm. We also paid attention to the achieved FPR_w values. Different versions are picked so that the amount of false positives is about the same for all algorithms, making comparisons possible.

Results for algorithm II are presented in Tables 12 and 13. RWSH v_3 reaches $\mathrm{Se_{AO}}$ 1st of 0.53 with $\mathrm{FPR_w}$ of 0.98, clearly outperforming both RWL and RWLF. Table 14 lists problematic patients for this algorithm.

Table 12: Detection results for algorithm II. This table presents Se_{AO} values.

	RWL			RWSH			RWLF		
Ident.	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_3$	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_3$	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_3$
case016	0	0	0	0	0.33	0	0	0.33	0.33
case035	0	0	0	0	0	0	0	0	0
case098	0	0	0	0	0	0	0	0	0
case101	0.20	0.76	0.32	0.08	0.28	0.20	0.24	0.84	0.60
case102	0	1	0	0	1	0	0	0	0
case103	0	0.10	0	0	0	0	0	0.20	0
case104	0	0	0	0	0	0	0	0	0
case106	0	1	0	1	1	1	0	1	0
case109	0	1	0	1	1	1	0	0	0
case119	1	1	1	0	1	1	0	0	0
case122	0	0.67	0	0.33	0.67	0.67	0.33	1	1
case126	0	0.80	0.30	0	0.70	0.50	0.40	0.90	0.90
case130	0	0	0	0	1	0	0	0	0
case144	0	0.50	0	0	0.40	0.10	0	1	0.90
case156	0	0	0	0	0	0	0	0	0
case157	0	0	0	0	0	0	0	0	0
case159	0	0.67	0	0.33	0.33	0.33	0	0.33	0
case177	0.17	0.33	0.25	0.08	0.33	0.25	0.17	0.58	0.25
case178	0	0.60	0	0	0.20	0	0	0	0
Thresholds:									
	88	65	80	86	68	75	52	30	40

6.4 Performance of algorithm III

The convex hull algorithm requires a slightly different approach. Since the algorithm does not produce a trend like the other two algorithms, we cannot use the same analysis methods. In this algorithm, there are two parameters that can be adjusted, $t_{\min}$ and $d_{\min}$.

We will first have to find out appropriate values for $t_{\min}$. Based on experience, we selected values of 5, 7, and 10 steps. Selecting $d_{\min}$ was done using same principles as with the other algorithms. After running the algorithm for seizure recordings, we inspected the values d obtained during or near to seizures. Based on this experiment, we completed runs for the full database using parameters listed in Table 15.

Because of the computational complexity of this algorithm, we could not afford to perform similar performance assessment as with the other algorithms. Instead, we selected only a specific and a sensitive version for each $t_{\min}$ value. Results are presented in Tables 16 and 17. Problematic patients for this method are listed in Table 18.

The largest effect of different $t_{\min}$ values is seen in case130 where $t_{\min} = 10$ does not yield a detection while the other designs do. In the other recordings the effect

Table 13: Statistics for algorithm II.

		$\mathbf{FPR_w}$	$\mathbf{FPR_{pa}}$	$\mathbf{Se_{AO}}$ **1st**	$\mathbf{Se_{AO}}$ **pt**	$\mathbf{Se_{AO}}$ **all szs**	$\mathbf{Se_{AO}}$ **pt szs**	$\mathbf{Sp_{Int,pa}}$	$\mathbf{Sp_{Int,w}}$	$\mathbf{Se_{Int}}$
RWL	v_1	0.84	0.93	0.16	0.16	0.06	0.07	0.99	0.99	0.01
	v_2	1.72	1.97	0.53	0.63	0.35	0.44	0.99	0.99	0.11
	v_3	1.04	1.16	0.21	0.21	0.11	0.10	0.99	0.99	0.03
RWSH	v_1	0.69	0.75	0.16	0.32	0.05	0.15	1.00	1.00	0.03
	v_2	1.25	1.35	0.63	0.68	0.23	0.43	0.99	0.99	0.14
	v_3	0.98	1.06	0.53	0.47	0.14	0.27	0.99	0.99	0.08
RWLF	v_1	0.30	0.63	0.21	0.21	0.09	0.06	1.00	1.00	0.03
	v_2	2.01	3.29	0.47	0.47	0.40	0.33	0.98	0.97	0.15
	v_3	0.82	1.55	0.37	0.32	0.29	0.21	0.99	0.99	0.08

Table 14: Detection problems with algorithm II.

Group	**Identifiers**
Seizure patients	case035, case098, case102, case103, case104, case130, case156, case157, case178
Non-seizure patients	case004, case006, case015, case017, case021, case039, case046, case047 case075, case083, case090, case119 case128, case168, case179, case180, vic005

of $t_{\min}$ is rather small. As expected, the algorithm becomes more specific as $d_{\min}$ is increased. This algorithm has a slightly lower integral specificity measures than the other two algorithms. This results from the nature of this algorithm; it attempts to capture the whole epoch that does not fit the background.

Table 15: Set-ups used for algorithm III.

$t_{\min}$	$d_{\min}$
5	0.33, 0.53, 0.77, 1.1, 1.2
7	0.5, 0.75, 1.2, 1.35
10	0.5, 0.75, 1.1, 1.35

Table 16: Results for selected specific (v_1) and sensitive (v_2) parameter combinations of algorithm III. This table presents Se_{AO} values.

	$t_{\min}$= 5		$t_{\min}$= 7		$t_{\min}$= 10	
Ident.	$\mathbf{v_1}$	$\mathbf{v_2}$	$\mathbf{v_1}$	$\mathbf{v_2}$	$\mathbf{v_1}$	$\mathbf{v_2}$
case016	0.33	0.67	0.33	0.67	0.33	1
case035	0	0.15	0	0.15	0.08	0.15
case098	0.33	0.33	0.33	0.33	0.33	0.67
case101	0.36	0.80	0.36	0.80	0.48	0.84
case102	0	1	0	1	0	1
case103	0	0.10	0	0.30	0	0.50
case104	0	0	0	0	0	0
case106	1	1	1	1	1	1
case109	1	1	1	1	1	1
case119	1	1	0	1	0	1
case122	1	1	1	1	1	1
case126	0.90	1	0.90	1	1	1
case130	1	1	1	1	0	0
case144	0.90	0.90	1	0.90	0.90	0.90
case156	0	0	0	0	0	0
case157	0.03	0.10	0.03	0.19	0.03	0.32
case159	0	0	0	0	0	0
case177	0.33	0.42	0.33	0.42	0.42	0.42
case178	0	0	0	0	0	0
$d_{\min}$						
	1.2	0.77	1.35	0.75	1.35	0.75

Table 17: Statistics for algorithm III.

		$\mathbf{FPR_w}$	$\mathbf{FPR_{pa}}$	$\mathbf{Se_{AO}}$ **1st**	$\mathbf{Se_{AO}}$ **pt**	$\mathbf{Se_{AO}}$ **all szs**	$\mathbf{Se_{AO}}$ **pt szs**	$\mathbf{Sp_{Int,pa}}$	$\mathbf{Sp_{Int,w}}$	$\mathbf{Se_{Int}}$
$t_{\min} = 5$	v_1	0.62	1.06	0.47	0.63	0.30	0.43	0.99	0.99	0.24
	v_2	1.82	2.58	0.63	0.79	0.45	0.55	0.97	0.97	0.33
$t_{\min} = 7$	v_1	0.54	0.83	0.47	0.58	0.30	0.38	0.99	0.98	0.23
	v_2	2.05	2.85	0.63	0.79	0.48	0.57	0.96	0.96	0.37
$t_{\min} = 10$	v_1	0.63	0.92	0.42	0.58	0.33	0.35	0.98	0.98	0.22
	v_2	1.99	2.67	0.58	0.74	0.54	0.57	0.95	0.94	0.40

Table 18: Detection problems with algorithm III.

Group	**Identifiers**
Seizure patients	case035, case104, case156 case157, case159 case178
Non-seizure patients	case006, case007, case013, case015, case017, case021, case085, case086, case090, case107, case116, case140, case142, case171, case173, case180, vic005

6.5 Comparison between the algorithms

In this subsection, we will highlight differences in how well the introduced algorithms perform. We will address both the problem of detecting seizures and the problem of false positive detections.

6.5.1 Detecting seizure patients

In the following comparisons, we consider both the specific and the sensitive variants of each algorithm. If we look at the problem from the clinical standpoint, paying attention mostly to the measure 1st ES, we cannot see much difference between the algorithms. By this measure, all algorithms identify correctly about 60% of the seizure patients. However, not all algorithms detect the same patients. Algorithm I performs poorly for case102, while other methods work well in that recording. Algorithm III outperforms the others in case103. In case177, algorithm II does not give a detection in Se_{AO} 1st sense. In most recordings, the results do not differ between algorithms.

Looking at the measure Se_{AO} pt, we can see that algorithm III performs the best. Even the specific version detects on average 60% of seizures of each patient. The sensitive version of algorithm III gives detections in Se_{AO} sense for all patients except for case104, case156, case159, and case178. Algorithm I does not provide additional information in these recordings, but algorithm II does detect seizures in case159 and in case178. To summarize, in Se_{AO} sense none of the algorithms detect seizures of recordings case104 and case156.

A drastic difference in performance is seen if we look at Se_{Int}. The specific version of algorithm III reaches 0.23, while for algorithms I and II we get 0.05 and 0.03, respectively. Similar performance gap is seen with the sensitive versions.

A summary of overall seizure detection potential is presented in Table 19. Potentially useful algorithms for each patient are marked by a cross. An algorithm is considered useful if detects most seizures of the patient in Se_{AO} sense.

If we look at the detection results as a whole, we can say that case035, case103,

Table 19: Summary of seizure detection potential of different methods. A cross indicates that the method has potential in detecting seizures of the corresponding patient.

	Identifier	**case016**	**case035**	**case098**	**case101**	**case102**	**case103**	**case104**	**case106**	**case109**	**case119**	**case122**	**case126**	**case130**	**case144**	**case156**	**case157**	**case159**	**case177**	**case178**
Algorithm	I	x							x	x	x	x		x						
	II							x	x	x	x	x								
	III	x		x	x			x	x	x	x	x	x	x						

case104, case156, case157, case159, case177, and case178 cannot be reliably detected by any of the methods presented. These recordings fall into three main categories:

1. SE recordings
2. Recordings with abundant spikes
3. Recordings with seizures that have no evolution in feature values

Case035 and case177 are near to SE, which means that ictal activity is constantly present and there might not be that much evolution left. In visual analysis only minuscule changes in feature values were seen during seizures. According to preliminary results with the spike detector developed in parallel with this project, we can see that case103, case157, case159, and case177 have a lot of spikes. Hence, they do not fall into the group of seizures the methods presented in this thesis are designed for. The rest belong to the third category. In seizures of case104, case156, and case178 there are no notable changes in the feature values. For these seizures, another strategy must be used. It should be noted that case156 had been cropped in hospital and there was only little data outside the annotated seizures. This prevented adaptive methods from detecting the seizures.

To conclude, we note that algorithm III performs generally the best out of the three proposed methods. There are, however, occasions when the other two methods provide useful additional information.

6.5.2 False positive detections

Table 20 presents data from analysis conducted with the identified problematic non-seizure recordings. We can see that there are four main causes for false positive detections:

1. Abundant EMG
2. Other artefacts
3. Naive detection logic
4. Detections arising from non-epileptiform EEG patterns.

An example of an EMG burst shown in Fig. 25. Corresponding feature values are shown in Fig. 26. Recordings with intermittent EMG bursts yield a lot of detections, especially with algorithm I.

Detection of artefacts was done in a rudimentary way and certain kinds of artefacts were, despite the rather strict criteria, passed on to the algorithms. This was especially problematic for adaptive methods. In particular, algorithm II was found to be very prone to artefacts.

We used a simple method to generate detections: if indicator value was above given threshold for more than 10 s, a detection was made. A more sophisticated method, merging together short nearby detections in different channels, would have

Table 20: Summary of patients with most false positive detections. Causes of false detections are listed for each algorithm.

	Source	**Reason for false positive detections**		
Identifier	**algorithm**	**I**	**II**	**III**
case004	II	–	Artefacts	–
case006	*ALL*	*Further expert review needed*		
case007	III	EMG	–	EMG
case013	III	–	–	1 EMG detection
case015	All	EMG	EMG	EMG
case017	All	EMG	–	1 EMG detection
case021	II, III	EMG	EMG	EMG
case039	II	EMG	EMG and artefacts	–
case044	I	EMG and detection logic	–	–
case045	I	EMG	–	1 EMG detection
case046	II	–	Detection logic	–
case047	II		Artefacts	
case050	I	EMG	–	
case056	I	High IF	–	–
case066	I	Detection logic	–	–
case075	II	–	EMG and artefacts	1 EMG detection
case083	II	–	Artefacts	–
case085	III	–	Detection logic	1 long detection
case086	III	–	–	EMG and artefacts
case090	All	EMG	EMG	EMG
case096	I	Drop in IF	Detection logic	–
case107	III	EMG		EMG
case116	I, III	EMG	–	EMG
case119	II		EMG and slow waves	–
case128	II	–	Artefacts	–
case140	I, III	EMG and artefacts	–	1 EMG detection
case142	I, III	EMG and artefacts	Artefacts	EMG
case146	*I*	*Modified burst–suppression*		
case168	II	–	Artefacts	–
case171	III	EMG	–	EMG
case173	III	EMG	–	EMG
case179	II	–	Detection logic	–
case180	All	EMG	EMG	EMG
vic005	II, III	–	Detection logic	β EEG

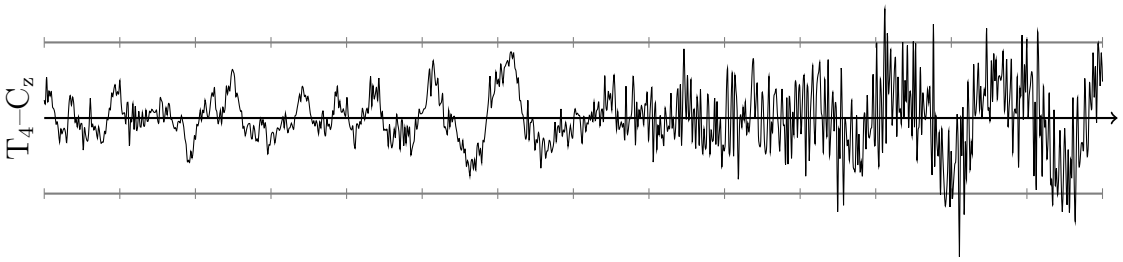

Figure 25: Example of EEG with a burst of EMG. The black line marks baseline, grey lines show ±50 µV values, and tick marks are printed every second.

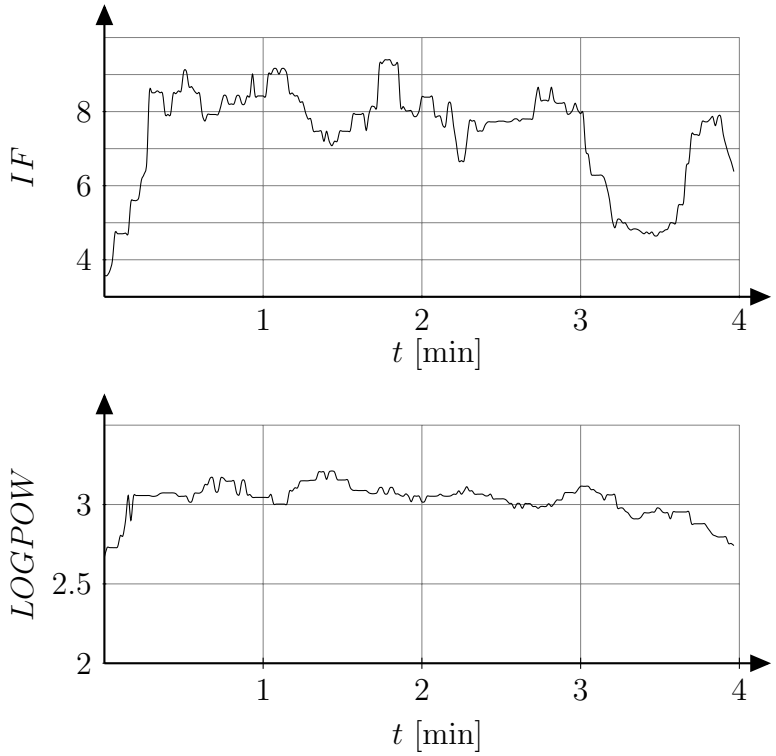

Figure 26: Feature tracings showing an EMG burst. Data starts at the same instance as the EEG tracing of Fig. 25. On-set of the burst, at about 10 s from the start, is seen as rapid increase in both feature values.

enhanced performance numbers. Adaptive methods were also found to be initialized with too little data in the buffer. This caused several detections, especially when using algorithm II, at the starts of the recordings.

One EEG pattern was discovered to cause several false positive detections. Fig. 27 shows an example of a modified burst–suppression pattern. During bursts, there is a strong β component. Faster activity is suppressed intermittently and a higher amplitude δ activity is present. This pattern causes large fluctuations in the feature values, leading to several detections. See Fig. 28 for corresponding feature traces.

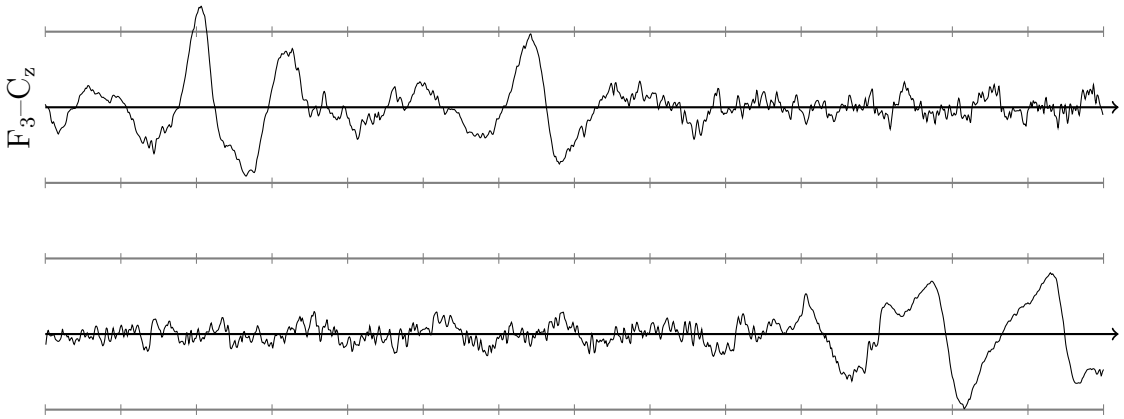

Figure 27: Modified burst-suppression pattern. This recording shows a rare pattern with alternating δ activity (suppression) and bursts of β activity. This 28 s example shows one cycle. The black line marks baseline, grey lines show $\pm 50\,\mu V$ values, and tick marks are printed every second.

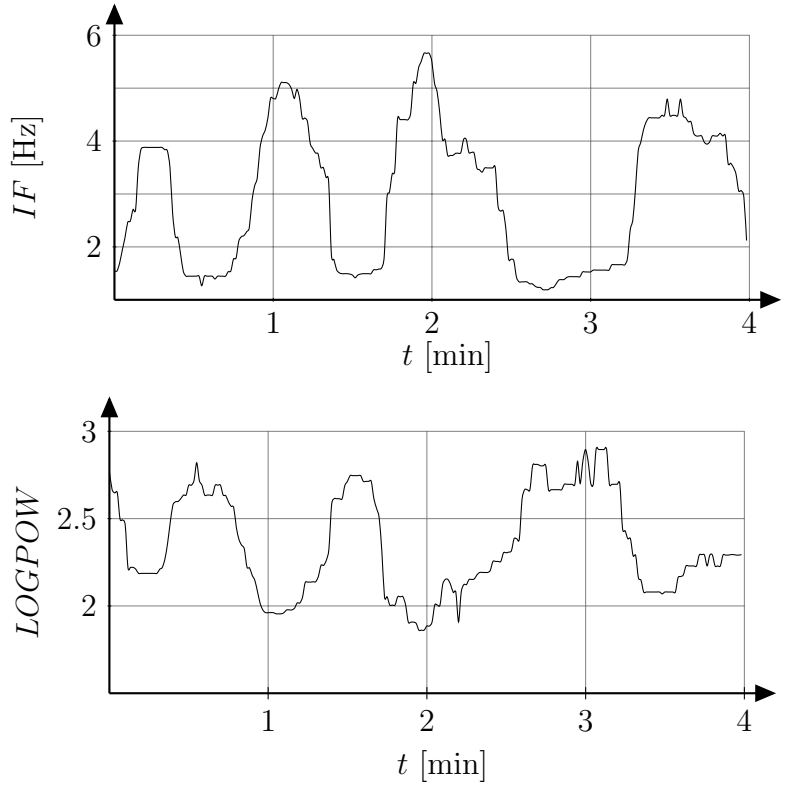

Figure 28: Feature tracing showing several cycles of modified burst–suppression pattern. Data starts at the same instance as the EEG tracing of Fig. 27. Bursts are associated with rapid increase in IF and decrease in $LOGPOW$.

7 Discussion

The methods presented in this thesis provide a novel way of investigating seizure activity in EEG. Instead of looking at snapshots of EEG or at absolute values of signal features, we approach the problem using the paradigm that in certain cases seizures cause systematic changes in signal properties. Absolute feature values are also important; most of the current methods use them and they seem to perform well in some patient populations. Our methods focus on dynamic changes in feature time courses.

During the study, several compromises had to be made. Despite the difficulties, eventually a protocol for assessing the performance of the three proposed methods was laid out. Analyses conducted in this study mark the way for future research. The proposed methods proved to hold potential for seizure detection.

7.1 Study

Throughout the study we have been cautious about reporting statistical variables and about drawing conclusions based solely on numbers. This has been an informed decision. We justify this approach by two arguments.

There is no single widely accepted performance measure for a seizure detection algorithm. Some attempts have been made, e.g., by Wilson [61], but in author's opinion the best overall measure is still escaping the researchers. One factor complicating the design of such a measure is the lack of proper specifications. It is easy to define how a perfect algorithm would work, but defining what is *rather* good or comparing two *almost* perfect algorithms is a difficult task. In author's opinion, the field lacks a common way of reporting results and describing performance. This leads to ambiguous reports and difficulties in making comparisons between methods.

Providing a dozen performance measures might be an accurate description of how well the algorithm performs, at least for researches with a background in science and technology. But we cannot expect medical staff to be specialists in statistics. Designing a widely accepted and simple performance measure might be a utopia but could also be worth striving for.

Another source of inaccuracy lies in what we consider as the ground truth. For this study, we had obtained the opinion of one seasoned expert, Dr. Bryan Young. Even though our consultant is a recognized specialist, we cannot guarantee that there are no mistakes in the annotations. And even if there would not be any bigger mistakes, another neurologist could have a different opinion regarding the exact onsets and off-sets of seizures. The best way to overcome this challenge would be to obtain dictations from other specialists and then to combine them. So far this has been limited by resources available for the project, but there are plans to obtain more dictations from other specialists.

In the presence of these two factors, we decided not to emphasize too much the statistical numbers and, instead, tried to look at the methods in a more general way.

Statistical optimization methods are very useful when designing algorithms. In this study, we wanted to investigate the potential of three proposed algorithms. This

led to a situation where we could not afford optimizing any of them in a statistically satisfying way. We made an informed decision that it is more useful to try and understand the general potential of the proposed methods than to blindly pick one of them and attempt at tuning it to its full potential.

While the absence of statistical methods could be seen as a scientific shortcoming in the study, it has allowed us to gain a more thorough understanding of the characteristics of the proposed methods. Such experience is invaluable when taking the next steps in the development process.

7.2 Characteristics of the designed algorithms

The most intuitive algorithm is the path length algorithm. When evaluating the method, certain problems were discovered. First and foremost, there were a lot of detections because of EMG. Second problem is that the method is not scale-invariant. In some recordings, we could notice a clear rise in path length values during seizures, proving the idea of the algorithm correct, but the total obtained path length values were far too low to yield detections.

Third problem with the path length method was its poor specificity. We wanted to sensitize the method for seizure-related events and added the weighting factor. However, we found some seizures in the database that do show evolution in feature values but not in the direction we used. Results could be somewhat different if the direction of the weighting vector $\vec{p}$ was optimized or if a different weighting scheme was implemented.

Overall, the path length algorithm was found to work somewhat well. In most cases its performance numbers were not as good as those of the other algorithms but it showed consistently acceptable results. The most appealing feature of the path length algorithm is its simplicity.

During evolution, the path traced by the features seems be drawn from a deterministic process. When there is no seizure activity, the path looks like a collection of randomly placed edges. The random walk algorithm was designed to quantify this change. We studied auto-regressive processes and the theory of random walk. Statistical tests were applied to understand if there really was a transition from a random or auto-regressive process to another process at seizure on-set. It was quickly established that such methods are not robust enough for our application and that a new method should be designed.

The random walk algorithm quantifies changes in feature values rather well. Like the path length algorithm, it produces a trend. The output value is proportional to the amount how much the current step differs from the past steps.

In the analyses reported here, we have seen that this method has a rather good performance. In several cases it seems to outperform the path length algorithm. While the scaling problem is less pronounced with this method, some seizures were still missed because of it. A bit surprisingly, this method seems to be the most robust against EMG. The current implementation was perhaps a bit faulty because, according to the analysis of false positive detections, this method is the one most susceptible to artefacts. It is hard to estimate how much the performance could

be improved by re-programming the routines and how much of the susceptibility is due to the actual method. It should be noted that the random walk algorithm was the only one that did not have any kind of directional preference for evolution implemented; the path length algorithm made use of weighting vector $\vec{p}$ and convex hull algorithm used the sector of accepted angles ν relative to the center-of-mass.

The last of the proposed methods, the convex hull algorithm, was based on the visual remark that during seizures the feature values often enter new areas in the feature space. As we know, these deviations can occur in various scales. Therefore, it would be advantageous to detect all such changes. We started bounding the background in rectangles and ellipses but quickly abandoned such shapes as they were too unstable. The notion of convex hull proved to be a good choice for this task.

This method proved to be the most complex one to implement. There are several special cases that need to be addressed. The implementation was tested but we can be fairly certain that the current implementation still has limitations. These special cases most likely have caused an excessive amount of long false positive detections.

The convex hull algorithm achieved its goal: it is capable of detecting events on all scales, making it the most sensitive one of the proposed methods. Performance numbers are in favour of developing this method further.

7.3 Main findings of the study

Algorithm III, the convex hull algorithm, was seen to have the best overall performance. However, there were recordings in which the other methods were more accurate. The convex hull algorithm is potentially useful in 11 out of 19 seizure patients included in the study. The developed methods detected almost all evolution seizures.

Most of the seizure patients that remained out of the scope of the developed methods have a lot of spikes. They are in the domain of the spike detector. Because the spike detector is still under development, we cannot provide any exact performance numbers here. Suffice it to say that after running spike detector, most likely only two patients will remain undetected.

According to visual review, the seizures in recording case104 seem to have evolutionary characteristics but the evolution occurs in such a low frequency that it is not captured by the features used in this study. Case178 is similar to case104.

At the current stage, all methods produce too many false positive detections. EMG was found to be a major contributor to those detections. While at first glance it might seem a problem easily dealt with, there are some complicating factors. Defining which patterns count as EMG activity is already a hard task as such. But most importantly, we do not want to lose the capability of detecting convulsive seizures. During convulsions there is typically a high amount of EMG in EEG.

In this study, we applied a very straightforward detection logic: if indicator value is above a given threshold for more longer than 10 s, a detection is made. There was no minimum time that detection should be off before it can be raised again. Fine-tuning the detection logic could improve the results. This type of problem

was seen with the path length algorithm. Implementing a proper way of handling multichannel data could also improve the performance numbers.

Artefacts proved to be a real nuisance. Even though we applied rather strict criteria to exclude artefactual signal epochs, some of them still got through to the algorithms. Especially the random walk algorithm was found to be sensitive to artefacts. We remind the reader that in parallel with this study, a new method for detecting movement artefacts was being developed. We hope that in the future a designated algorithm can handle artefact detection reliably.

With the adaptive methods, random walk and convex hull, a different detection logic problem was encountered. When the buffer was not full, e.g., at the beginning of each recording, the outputs from these algorithms were noisy. On the one hand, we should always ensure that the buffer is full before any computations are made. On the other hand, because of cropped recordings we would then have lost several seizures. This issue is specific to this stage of development and should not be a problem if these methods are implemented for on-line monitoring.

We reported only one EEG pattern that causes excessive amount of false positive detections. Keeping in mind that there is more than 150 d of data, this is not a bad result. However, after taking care of the current problems, like EMG, we will most likely find more EEG patterns that need special attention.

7.4 Guidelines for future development

In the future, we will try to combine the good sides of each algorithm. One possible scheme is to use the convex hull algorithm for making elementary detections. After making such a detection, the following epoch is subject to further testing. Tests could comprise of path length, spike rate, distance, or random walk, to name a few possibilities. Using the convex hull algorithm would help reduce false positives and would also solve the scaling problem.

It is evident that in the next step we will have to start integrating evolution detector, spike detector, and artefact detector into one package. So far each method has been developed separately. Only when all these methods are put together can we get realistic performance numbers of the seizure detection algorithm.

At the same time we need to continue to develop our database. The more data there is available for development, the better generalization properties we can expect. Getting another neurologist to annotate the recordings can be costly, but it would increase the creditability of the database. Having a large and well-maintained database is a valuable asset as such.

On a side note, we should keep in mind also the lack of a single proper performance measure for this type of studies. If we could agree with the experts on how they define good performance and translate it into a performance indicator, it would be useful for the entire community.

8 Conclusions

This thesis addressed the problem of detecting seizures that have an evolutionary pattern. Three methods were proposed and implemented. Their performance was assessed using both statistical methods and visual review. We found that the convex hull algorithm reached the best overall performance level. The idea of that method is to enclose background activity in a convex hull. When the hull is breached, further tests are conducted to deduce whether the event is related to seizure activity or not.

We got evidence that the convex hull method alone, however, does not suffice. It needs to be developed further, possibly by integrating other methods into it. In the next phase, we must also address false positive detections with more rigour than we did in this study.

While the results obtained in this study are encouraging, it is too early to compare them with those of published methods. Before an overall assessment can be made, our method needs to be fine-tuned and must incorporate a proper artefact detector and a spike detector.

References

[1] J Engel, T Pedley, and J Aicardi. *Epilepsy: A comprehensive textbook (Vol. 1)*. Lippincott Williams & Wilkins, Philadelphia, 2nd edition, 2007.

[2] I Grant and P Andrews. ABC of intensive care: Neurological support. *BMJ*, 319(7202):110–113, 1999.

[3] E Ikonen. *Muotoiluhaasteena EEG-sensorisetin käyttö pitkäaikaisseurannassa*. BSc thesis, Aalto University, 2010.

[4] A Savelainen. *Movement artifact detection from electroencephalogram utilizing accelerometer*. MSc thesis, Aalto University, 2011.

[5] L Haas. Neurological stamp. *J. Neurol. Neurosur. Ps.*, 74(5):653, 2003.

[6] B Nielsen. *Towards 24-7 Brain Mapping Technology*. MSc thesis, Technical University of Denmark, 2009.

[7] H Jasper. The ten-twenty electrode system of the International Federation. *Electroen. Clin. Neuro.*, 10:371–375, 1958.

[8] J Malmivuo and R Plonsey. *Bioelectromagnetism—Principles and applications of bioelectric and biomagnetic fields*. Oxford University Press, New York, 1st edition, 1995.

[9] B Fisch. *EEG Primer: Basic principles of digital and analog EEG*. Elsevier, Amsterdam, 3rd edition, 1999.

[10] L Hirsch and R Brenner. *Atlas of EEG in Critical Care*. John Wiley & Sons, 1st edition, 2010.

[11] B Young, R McLachlan, J Kreeft, and J Demelo. An electroencephalographic classification for coma. *Can. J. Neurol. Sci.*, 24(4):320–325, 1997.

[12] H Viertiö-Oja, V Maja, M Särkelä, P Talja, N Tenkanen, H Tolvanen-Laakso, M Paloheimo, A Vakkuri, A Yli-Hankala, and P Meriläinen. Description of the Entropy algorithm as applied in the Datex-Ohmeda S/5 Entropy Module. *Acta Anaesth. Scand.*, 48(2):154–161, 2004.

[13] P Vespa, V Nenov, and M Nuwer. Continuous EEG monitoring in the intensive care unit: early findings and clinical efficacy. *J. Clin. Neurophysiol.*, 16(1):1–13, 1999.

[14] D Friedman, J Claassen, and L Hirsch. Continuous electroencephalogram monitoring in the intensive care unit. *Anesth. Analg.*, 109(2):506–523, 2009.

[15] M Mirski and P Varelas. Seizures and status epilepticus in the critically ill. *Crit. Care Clin.*, 24:115–147, 2008.

[16] J Claassen, S Mayer, R Kowalski, R Emerson, and L Hirsch. Detection of electrographic seizures with continuous EEG monitoring in critically ill patients. *Neurology*, 62(10):1743–1748, 2004.

[17] M Scheuer. Continuous EEG monitoring in the intensive care unit. *Epilepsia*, 43 Suppl 3:114–127, 2002.

[18] B Young, J Ives, M Chapman, and S Mirsattari. A comparison of subdermal wire electrodes with collodion-applied disk electrodes in long-term EEG recordings in ICU. *Clin. Neurophysiol.*, 117(6):1376–1379, 2006.

[19] R Agarwal and J Gotman. Long-term EEG compression for intensive-care settings. *IEEE Eng. Med. Biol.*, 20(5):23–29, 2001.

[20] Seizure. Merriam-Webster Online Dictionary, http://www.merriam-webster.com/medical/seizure, visited 29.4.2011.

[21] W Hauser, J Annegers, and L Kurland. Prevalence of epilepsy in Rochester, Minnesota: 1940–1980. *Epilepsia*, 32(4):429–445, 1991.

[22] B Purcell, A Gaitatzis, J Sander, and A Majeed. Epilepsy prevalence and prescribing patterns in England and Wales. *Health Stat. Quart.*, 15:23–30, 2002.

[23] B Young, K Jordan, and G Doig. An assessment of nonconvulsive seizures in the intensive care unit using continuous EEG monitoring: an investigation of variables associated with mortality. *Neurology*, 47(1):83–89, 1996.

[24] R Alroughani, M Javidan, A Qasem, and N Alotaibi. Non-convulsive status epilepticus; the rate of occurrence in a general hospital. *Seizure*, 18(1):38–42, 2009.

[25] S Gupta and A Parihar. Seizures in the intensive care unit. *JK Sci.*, 2(2):81–87, 2000.

[26] K Jordan. Neurophysiologic monitoring in the neuroscience intensive care unit. *Neurol. Clin.*, 13(3):579–626, 1995.

[27] R DeLorenzo, E Waterhouse, R Towne, J Boggs, D Ko, G DeLorenzo, A Brown, and L Garnett. Persistent nonconvulsive status epilepticus after the control of convulsive status epilepticus. *Epilepsia*, 39(8):833–840, 1998.

[28] P Vespa, M Nuwer, V Nenov, E Ronne-Engstrom, D Hovda, M Bergsneider, D Kelly, N Martin, and D Becker. Increased incidence and impact of non-convulsive and convulsive seizures after traumatic brain injury as detected by continuous electroencephalographic monitoring. *J. Neurosurg.*, 91(5):750–756, 1999.

[29] P Vespa, K O'Phelan, M Shah, J Mirabelli, S Starkman, C Kidwell, J Saver, M Nuwer, J Frazee, D McArthur, and N Martin. Acute seizures after intracerebral hemorrhage. *Neurology*, 60:1441–1446, 2003.

[30] J Pandian, G Cascino, E So, E Manno, and J Fulgham. Digital video-electroencephalographic monitoring in the neurological-neurosurgical intensive care unit. *Arch. Neurol.*, 61:1090–1094, 2004.

[31] N Jetté, J Claassen, R Emerson, and L Hirsch. Frequency and predictors of nonconvulsive seizures during continuous electroencephalographic monitoring in critically ill children. *Arch. Neurol.*, 63:1750–1755, 2006.

[32] J Claassen, N Jetté, F Chum, R Green, M Schmidt, H Choi, J Jirsch, J Frontera, E Connolly, R Emerson, S Mayer, and L Hirsch. Electrographic seizures and periodic discharges after intracerebral hemorrhage. *Neurology*, 69(13):1356–1365, 2007.

[33] M Oddo, E Carrera, J Claassen, S Mayer, and L Hirsch. Continuous electroencephalography in the medical intensive care unit. *Crit. Care Med.*, 37(6):2051–2056, 2009.

[34] D Chong and L Hirsch. Which EEG patterns warrant treatment in the critically ill? Reviewing the evidence for treatment of periodic epileptiform discharges and related patterns. *J. Clin. Neurophysiol.*, 22(2):79–91, 2005.

[35] S Wilson, M Scheuer, C Plummer, B Young, and S Pacia. Seizure detection: correlation of human experts. *Clin. Neurophysiol.*, 114(11):2156–2164, 2003.

[36] Epileptiform. Merriam-Webster Online Dictionary, http://www.merriam-webster.com/medical/epileptiform, visited 29.4.2011.

[37] J Gotman. Automatic detection of epileptic seizures. In *Handbook of clinical neurophysiology, vol. 3*, pages 155–165. Elsevier, 2004.

[38] J Gotman. Automatic detection of seizures and spikes. *J. Clin. Neurophysiol.*, 16(2):130–140, 1999.

[39] J Gabor, R Leach, and F Dowla. Automated seizure detection using a self-organizing neural network. *Electroen. Clin. Neuro.*, 99(3):257–266, 1996.

[40] H Firpi. Epileptic seizure detection using genetically programmed artificial features. *IEEE Trans. Biomed. Eng.*, 898(2):212–224, 2007.

[41] A Zandi, G Dumont, M Javidan, R Tafreshi, B MacLeod, C Ries, and E Puil. A novel wavelet-based index to detect epileptic seizures using scalp EEG signals. In *IEEE EMBS 2008*, pages 919–922.

[42] G Tezel and Y Ozbay. A new approach for epileptic seizure detection using adaptive neural network. *Expert Syst. Appl.*, 36(1):172–180, 2009.

[43] L Guo, D Rivero, J Dorado, J Rabuñal, and A Pazos. Automatic epileptic seizure detection in EEG based on line length feature and artificial neural networks. *J. Neurosci. Meth.*, 191:101–109, 2010.

[44] A Zandi, M Javidan, G Dumont, and R Tafreshi. Automated real-time epileptic seizure detection in scalp EEG recordings using an algorithm based on wavelet packet transform. *IEEE Trans. Biomed. Eng.*, 57(7):1639–1651, 2010.

[45] M Scheuer and S Wilson. Data analysis for continuous EEG monitoring in the ICU: seeing the forest and the trees. *J. Clin. Neurophysiol.*, 21(5):353–378, 2004.

[46] A Shah, R Agarwal, J Carhuapoma, and J Loeb. Compressed EEG pattern analysis for critically ill neurological-neurosurgical patients. *Neurocrit. Care*, 5(2):124–133, 2006.

[47] J Gotman. Automatic recognition of epileptic seizures in the EEG. *Electroen. Clin. Neuro.*, 54(5):530–540, 1982.

[48] J Gotman. Automatic seizure detection: improvements and evaluation. *Electroen. Clin. Neuro.*, 76(4):317–324, 1990.

[49] J Gotman and H Qu. Improvement in seizure detection performance by automatic adaptation to the EEG of each patient. *Electroen. Clin. Neuro.*, 86(2):79–87, 1993.

[50] S Wilson, M Scheuer, R Emerson, and J Gabor. Seizure detection: evaluation of the Reveal algorithm. *Clin. Neurophysiol.*, 115(10):2280–2291, 2004.

[51] M Särkelä, M Ermes, M van Gils, A Yli-Hankala, V Jäntti, and A Vakkuri. Quantification of epileptiform electroencephalographic activity during sevoflurane mask induction. *Anesthesiology*, 107(6):928–938, 2007.

[52] B Young, M Sharpe, M Savard, E Al Thenayan, L Norton, and C Davies-Schinkel. Seizure detection with a commercially available bedside EEG monitor and the subhairline montage. *Neurocrit. Care*, 11(3):411–416, 2009.

[53] B Kolls and A Husain. Assessment of hairline EEG as a screening tool for nonconvulsive status epilepticus. *Epilepsia*, 48(5):959–965, 2007.

[54] L Sörnmo and P Laguna. *Bioelectrical signal processing in cardiac and neurological applications*. Elsevier Acadmic Press, San Diego, 1st edition, 2005.

[55] M Le Van Quyen, J Foucher, J Lachaux, E Rodriguez, A Lutz, J Martinerie, and F Varela. Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony. *J. Neurosci. Meth.*, 111(2):83–98, 2001.

[56] J Dauwels, F Vialatte, T Musha, and A Cichocki. A comparative study of synchrony measures for the early diagnosis of Alzheimer's disease based on EEG. *Neuroimage*, 49(1):668–693, 2010.

[57] M Misiti, Y Misiti, G Oppenheim, and J Poggi. *Wavelet toolbox for use with Matlab*. The MathWorks Inc., Natick, 1st edition, 1996.

[58] S Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pat. Anal. Mach. Int.*, 11(7):674–693, 1989.

[59] J.-C. Pesquet, H. Krim, and H. Carfantan. Time-invariant orthonormal wavelet representations. *IEEE Trans. Signal Proces.*, 44(8):1964–1970, 1996.

[60] M Johansson. *The Hilbert transform*. MSc thesis, Växjö University.

[61] S Wilson. Algorithm architectures for patient dependent seizure detection. *Clin. Neurophysiol.*, 117(6):1204–1216, 2006.

[62] B Young. Personal communication, 2010.

[63] E Huigen. *Noise in biopotential recording using surface electrodes*. Msc thesis, Delft Technical University, 2000.

[64] C Barber, D Dobkin, and H Huhdanpää. The quickhull algorithm for convex hulls. *ACM Trans. Math. Soft.*, 22(4):469–483, 1996.

[65] Matlab 6.5, The Mathworks Inc., Natick, USA.