KU LEUVEN

ARENBERG DOCTORAL SCHOOL
Faculty of Engineering Science

Signal processing methods for home monitoring of people with epilepsy with a network of wearable EEG sensors

Jonathan Dan

Supervisors:
Prof. dr. ir. A. Bertrand
Dr. B. Vandendriessche
Prof. dr. md. W. Van Paesschen

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Electrical Engineering

November 2022

Signal processing methods for home monitoring of people with epilepsy with a network of wearable EEG sensors

Jonathan DAN

Examination committee:
Prof. dr. ir. Y. Willems, chair
Prof. dr. ir. A. Bertrand, supervisor
Dr. B. Vandendriessche, supervisor
Prof. dr. md. W. Van Paesschen, supervisor
Prof. dr. ir. S. Van Huffel
Prof. dr. ir. M. Kraft
Prof. dr. D. Atienza
(École Polytechnique Fédérale de Lausanne)
Dr. ir. J. Duun-Henriksen
(UNEEG)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Electrical Engineering

November 2022

Preface

In 2016, I joined Byteflies to develop wearable health sensors to transform health care from point-of-care interventions in a hospital to early diagnosis and monitoring in everyday life. Byteflies and several partners set up the SeizeIT consortium. The project aimed to improve the care and treatment of people with epilepsy. We were looking for solutions to record seizures in everyday life. Wearable accelerometry and electrocardiography, among others, were evaluated for their ability to record physiological features of seizures outside of the hospital. The project concluded that wearable electroencephalography is required to monitor all seizure types (including non-motor seizures, such as absence seizures). At the time, no wearable EEG met requirements for everyday use by people with epilepsy, so Byteflies set out to develop a wearable EEG sensor. My PhD thesis was conceived in this context. Together with Byteflies, UZ Leuven and KU Leuven, we formulated the challenges faced by wearable EEG. For this thesis, we decided to focus on the signal processing challenges related to wearable EEG. The project was supported by the Flemish government (VLAIO) and Byeflies through a Baekeland mandate.

And here we are today! In the four years since the beginning of my PhD the landscape has evolved. Some wearable sensors to monitor patients with generalised tonic-clonic seizures have become medical devices. Byteflies developed the Sensor Dot and measured many patients with different disorders ranging from cardiorespiratory infections such as COVID-19 to epilepsy. Other innovative companies have also developed new ways of monitoring epilepsy in the long term, e.g. subcutaneous EEG. This thesis intends to provide the community with some of the missing tools to solve the puzzle of long-term unobtrusive EEG measurement in people with epilepsy in their everyday life.

This project was made possible by the joint effort of all partners and collaborators. I would like to use the opportunity of this manuscript to thank wholeheartedly the entire team of epilepsy at UZ Leuven. The patients as the main users of this technology for inspiring this project for sharing their

experience and ideas. I thank you for your participation, trust and hope in this project. The EEG nurses who taught me how to record EEG and helped me collect data. The neurology residents who taught me how to read EEG, helped me annotate recordings and were always present to answer questions and collaborate on different aspects of this thesis. The neurology research assistants who guided the project from the initial stages of SeizeIT to where it stands today. As well as prof. dr. Wim Van Paesschen my co-supervisor, for his support and vision for the development of wearable EEG. I would also like to thank the whole team at Byteflies for training me first as a hands-on engineer in a highly dynamic start-up environment, then accompanying me along the journey of this PhD. I congratulate and thank all my colleagues for the huge progress that was made to provide new tools to the epilepsy community. I thank everyone for sharing the questions and problems you were facing, continuously feeding me with new ideas and challenges for my research. I would also like to personally thank Benjamin Vandendriessche, for being my mentor and teaching essentials and practicals of the biomedical engineer profession. You taught me to structure work, plan, develop, and validate biomedical signal processing algorithms. I thank you for being a challenging and constructive sparring partner as well as a meticulous supervisor. Finally, I would like to thank the whole biomed team of ESAT. The diversity of research topics and methods provides an eye-opening environment that is particularly fruitful for the development of personal research. I drew my biggest satisfaction from collaboration projects bringing together people with different expertise, such as those taking place in the DISPATCH Neuro-Sense and Neureka challenge. I would like to thank all my colleagues who shared their intelligence and advice throughout this project. I would particularly like to thank my supervisor Alexander Bertrand for guiding this project. Alex is one of the main brains behind the concept of EEG sensor networks. Your encyclopedic knowledge of signal processing algorithms and your didactic power were the main drivers behind the success of this project. I thank you for proposing the general thrust for the research and continuously enlightening me with new ideas and challenges.

A PhD project is accomplished within a community. I was lucky to be part of the epilepsy community. I would like to thank all patients and caregivers who explained the needs of the community and challenged the technical innovations we proposed. I am also grateful to the whole scientific community for the shared expertise and openness, as well as industrial players who have always worked in a spirit of collaboration to foster innovation.

I wish you a pleasant reading of this thesis.

Abstract

Epilepsy is one of the most common severely disabling brain conditions, affecting over 46 million people worldwide. It is characterised by recurring, sudden, excessive, synchronous electrical activity in neuronal networks that disrupt ongoing brain activity and causes clinical seizures. Diagnosis and follow-up of epilepsy typically relies on electroencephalogaphy (EEG). Clinical EEG is a non-invasive electrophysiological measurement technique recording the electrical field of the brain cortex from a number of scalp surface electrodes. It is typically acquired for a few dozens of minutes in a controlled condition in the clinic with devices that are ill-suited for long-term EEG-monitoring in daily life.

The development of portable EEG technology has taken advantage of the advent of silicon integrated circuit chips and the subsequent miniaturisation of electronics. However, the practical usability of many current EEG wearable devices is still limited. One major limitation relates to the trade-off between number of electrodes and wearability. More electrodes lead to larger electronics and more wires, which make the platform less user-friendly, and induce more wire-related artefacts. Conversely, the less electrodes, the lower the spatial resolution and the less tasks the system can be used for.

We propose a different platform to offer solutions to some of the current limitations associated with wearable EEG devices. This new platform will consist of a collection of wireless miniature EEG sensor units operating as a sensor network. Each EEG unit should incorporate electrodes, an amplifier, a wireless radio and a processing unit in a single package with a small scalp area footprint. The new topology eliminates the need for wires and will allow for a flexible, discreet, miniature, wearable system with as many EEG sensors as necessary for a particular application or patient. This new topology introduced new challenges which are investigated in this PhD thesis.

Specifically, this thesis presents signal processing methods for the analysis of EEG in a wireless EEG sensor network. It addresses the following main challenges.

1. Design algorithms for automatic analysis of epileptiform activity. To do this we developed a novel multi-channel EEG signal processing method for automated epileptiform event detection which is specifically designed to run on a microcontroller with minimal memory and processing power. It is based on a linear multi-channel filter that is precomputed offline in a data-driven fashion based on the spatial-temporal signature of the seizure and peak interference statistics. At run-time, the algorithm requires only standard linear filtering operations, which are cheap and efficient to compute, in particular on microcontrollers with a multiply-accumulate unit. It has been validated on multiple datasets and compared to existing state-of-the-art algorithms.

2. Design strategies for optimal EEG sensor selection. For this we propose a channel (or variable) selection algorithm for generalised eigenvalue problems such as those used in our epileptiform event detection algorithm. The method extends and generalises existing work on convex optimisation-based variable selection using semidefinite relaxations toward group-sparse variable selection using the $\ell_{1,\infty}$-norm. We comprehensively compared our method to state-of-the-art methods for sensor selection for spatio-temporal filter design in a simulated sensor network setting.

3. Provide guidelines for the design strategies of miniaturised EEG sensor networks. This was investigated by conducting a study on the limits of miniaturisation of an EEG sensor network by emulating different networks using high-density EEG recordings and analysing interictal spikes in the different simulations.

Results show that the epileptiform event detection algorithm performs on par with state-of-the-art detection algorithms at a much lower computational cost for the detection of absence seizures and interictal epileptiform discharges. The study on channel selection algorithms indicates which algorithm to select in function of computational constraints, number of channels to select, and the topology of the problem. It shows that both the proposed channel selection algorithm and a backward greedy selection method best approximate the optimal solution. The proposed algorithm is also more robust to failure cases. The study on limits of miniaturisation of a network of wireless EEG sensors showed that recording equipment should be specifically designed to measure the small signal power at a short inter-electrode distance by providing an input-referred noise of < 300 nV. It also showed that an inter-electrode distance of minimum

5 cm in a setup with a minimum of two EEG units is required to obtain near equivalent performance in interictal spike detection to standard EEG.

In summary, this PhD thesis introduces several new signal processing methods for wireless EEG sensor networks for monitoring of people with epilepsy. It contributes to the technological advancement required for the wider adoption of this technology.

Beknopte samenvatting

Epilepsie is een van de meest voorkomende ernstig invaliderende hersenaandoeningen, waaraan wereldwijd meer dan 46 miljoen mensen lijden. Het wordt gekenmerkt door terugkerende, plotselinge en overmatige synchrone elektrische activiteit in neuronale netwerken die de lopende hersenactiviteit verstoort en klinische aanvallen veroorzaakt. De diagnose en opvolging van epilepsie berusten op elektroencefalogafie (EEG). Klinische EEG is een niet-invasieve elektrofysiologische meettechniek die het elektrische opgewekt door de hersenen registreert via een aantal oppervlakte-elektroden op de hoofdhuid. Deze EEG data wordt meestal opgenomen in het ziekenhuis met apparatuur dat niet geschikt is voor langdurige EEG-monitoring in het dagelijks leven.

De ontwikkeling van draagbare EEG-technologie is mogelijk gemaakt door de komst van chips met geïntegreerde schakelingen en de daaropvolgende miniaturisering van elektronica. De bruikbaarheid van veel van de huidige draagbare EEG-apparaten is echter nog steeds beperkt. Een belangrijke beperking houdt verband met de afweging tussen het aantal elektroden en de draagbaarheid. Een groter aantal elektroden leidt tot grotere elektronica en meer draden, die het platform minder gebruiksvriendelijk maken en ook meer artefacten veroorzaken. Omgekeerd geldt dat hoe minder elektroden gebruikt worden, hoe lager de spatiale resolutie van het EEG wordt en voor hoe minder applicaties het systeem kan worden gebruikt. Wij stellen een alternatief platform voor om oplossingen te bieden voor enkele van de huidige problemen door meerdere draadloze miniatuur EEG-sensoreenheden te combineren en samen te laten werken in een gezamenlijk sensornetwerk. Elke EEG-eenheid bevat zijn eigen elektroden, een versterker, een draadloze radio en een verwerkingseenheid. Een dergelijk platform maakt draden overbodig en faciliteert een flexibel, discreet, en draagbaar systeem met zoveel EEG-sensoren als nodig voor een bepaalde toepassing of patiënt. Deze nieuwe technologie introduceert echter nieuwe uitdagingen die in dit proefschrift onderzocht worden.

Specifiek introduceert dit proefschrift signaalverwerkingsmethoden voor de

analyse van EEG in een draadloos EEG-sensornetwerk. Het behandelt de volgende belangrijke uitdagingen.

1. Ontwerp van algoritmen voor automatische analyse van epileptiforme activiteit. Om dit te doen hebben we een nieuwe meerkanaals EEG-signaalverwerkingsmethode ontwikkeld voor automatische epileptiforme detectie, speciaal ontworpen voor microcontrollers met een beperkt geheugen en rekenkracht. Deze methode is gebaseerd op een lineaire meerkanaals filter die vooraf wordt berekend op een datagedreven manier op basis van de spatio-temporele patronen van de epileptische aanval en de piekinterferentiestatistieken. Tijdens de uitvoering vereist het algoritme alleen standaard lineaire filterbewerkingen, die goedkoop en efficiënt te berekenen zijn, met name op microcontrollers met een vermenigvuldigings-accumulatie-eenheid. Deze methode wordt gevalideerd op meerdere datasets en vergeleken met bestaande state-of-the-art algoritmen.

2. Ontwerpstrategieën voor optimale EEG-sensorselectie. We hebben een nieuw kanaalselectie algoritme voor gegeneraliseerde eigenwaarde problemen ontworpen, wat gebruikt kan worden voor de detectie van epilepsie-gerelateerde patronen in het EEG. De methode breidt bestaand werk over variabelenselectie met behulp van semidefiniete relaxaties uit, waarbij nu ook selectie van *groepen* van variabelen mogelijk wordt. We hebben deze methode uitvoerig vergeleken met state-of-the-art methoden voor sensorselectie in een gesimuleerde sensornetwerksetting.

3. Richtlijnen voor ontwerpstrategieën van geminiaturiseerde EEG- sensornetwerken. Hier gaan we op zoek naar de grenzen van miniaturisatie van een EEG-sensornetwerk door sensoren van verschillende groottes te emuleren op basis van hoge-densiteit EEG-opnames en de analyse van interictale *spikes* in de verschillende simulaties.

Uit de resultaten blijkt dat het algoritme voor de detectie van epileptische patronen even goed presteert als de state of the art detectiealgoritmen, maar veel minder rekenkracht nodig heeft. De studie van algoritmen voor kanaalselectie geeft aanwijzingen over welk algoritme moet worden gekozen in functie van de computationele beperkingen, het aantal te selecteren kanalen en de topologie van het probleem. Het toont aan dat zowel het voorgestelde algoritme en een 'greedy' selectiemethode de optimale oplossing het best benaderen. Het voorgestelde algoritme is daarbij robuuster tegen falingen. De studie naar de grenzen van de miniaturisering van een netwerk van draadloze EEG-sensoren toonde aan dat de opnameapparatuur specifiek moet worden ontworpen om het kleine signaalvermogen bij een korte inter-elektrodeafstand te meten met een input-ruis van < 300 nV. Het toonde ook aan dat een inter-elektrode afstand

van minimaal 5 cm in een opstelling met minimaal twee EEG-eenheden nodig is om bijna gelijkwaardige prestaties in interictale *spike* detectie te verkrijgen als bij standaard EEG.

Samengevat introduceert dit proefschrift verschillende nieuwe signaalverwerkingsmethoden voor draadloze EEG-sensornetwerken voor het monitoren van mensen met epilepsie. Het draagt bij aan de technologische vooruitgang die nodig is voor een brede toepassing van deze technologie.

List of Abbreviations

BCI	brain computer interface.
BE	backward elimination.
CSP	common spatial patterns.
DC	direct current.
DF	degrees of freedom.
ECG	electrocardiography.
EEG	electroencephalograph, electroencephalography or electroencephalogram (*In the remainder of the text* EEG *is used interchangeably for the recording device and for the recorded signal*).
EMG	electromyography.
FFT	fast Fourier transform.
FIR	finite impulse response.
FPGA	field-programmable gate arrays.
FS	forward selection.
GEVc	generalised eigenvector.
GEVD	generalised eigenvalue decomposition.
GEVl	generalised eigenvalue.
GRQ	generalised Rayleigh quotient.
GUI	graphical user interface.
HD-EEG	high density electroencephalograph.
IC	integrated circuit.

IED	interictal epileptiform discharges.
IIR	infinite impulse response.
LDA	linear discriminant analysis.
LMEM	linear mixed-effect model.
MIMO	multiple-input multiple-output.
MISO	multiple-input single-output.
RMS	root mean square.
SE	standard error.
SNR	signal-to-noise.
SPIR	seizure to peak interference ratio.
SVM	support-vector machine.

List of Symbols

General mathematical notations

$.^T$	transpose
$.^{-1}$	inverse
$\mathcal{O}$	Big O notation
δ_{ij}	Kronecker-delta
$\mathbb{R}$	set of real numbers
$\mathbf{0}_N$	$N \times N$ matrix with zeros
$\mathbf{I}_N$	$N \times N$ identity matrix
$\mathbf{X} \succcurlyeq 0$	$\mathbf{X}$ is a positive semidefinite matrix
$\|\|.\|\|_{\max}$	max-norm of a matrix (maximal absolute value across all elements)
$\|\|\mathbf{x}\|\|_1$	ℓ_1-norm of a vector (sum of the absolute values)
$\|\|\mathbf{x}\|\|_2$	ℓ_2-norm of a vector (square root of the sum of squared elements)
$\|\|\mathbf{x}\|\|_\infty$	ℓ_∞-norm of a vector (maximal absolute value)
$\mathbf{A}$	matrix
$\mathbf{a}$	vector
$\mathrm{Tr}\,(.)$	trace operator
a	scalar
$E\{.\}$	expected value operator

x_{ij} element of matrix $\mathbf{X}$ on the i^{th} row and j^{th} column

$\mathbf{X} \otimes \mathbf{Y}$ Kronecker product of matrices $\mathbf{X}$ and $\mathbf{Y}$

Text specific symbols

$\mathbf{R}_n$ noise covariance matrix

$\mathbf{R}_s$ seizure covariance matrix

$\mathbf{w}$ vector containing filter coefficients

$\mathbf{x}(t), \mathbf{y}(t)$ N-dimensional vector containing the sample at time t collected at N EEG channels

C, N, M Number of EEG channels, sensors or variables

d inter-electrode distance

K Number of filter outputs

L Number of filter time lags (filter taps)

$o(t)$ single channel output of gilter at time t

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

The goal of this PhD thesis is to develop *signal processing methods to monitor people with epilepsy in a home environment with wearable EEG*. This introduction will outline the problem and background information required to read this dissertation. First the neurological condition of epilepsy is described in section 1.1. EEG as a recording technique and the resulting signal is described in section 1.2. Home EEG monitoring with ambulatory recording devices is introduced in section 1.3. An overview of standard practices in processing EEG signals (in particular for epilepsy) are presented in section 1.4. Following this, the research objectives are stated and an overview of the organisation of this dissertation are provided.

1.1 Epilepsy

Epilepsy is a brain disorder characterised by recurring, sudden, excessive, synchronous electrical activity in neuronal networks that disrupt ongoing brain activity and causes clinical manifestations known as seizures, i.e. stereotyped behaviours or sensations that can include alteration of consciousness, depending on the part of the brain that is affected. Seizure events are referred to as ictal, and periods between seizures are called interictal. The International League Against Epilepsy (ILAE) defines epilepsy operationally based on the following situations: 1) at least two unprovoked seizures occurring more than 24 hours apart; 2) at least one unprovoked seizure and probability of further seizures similar to the general recurrence risk (i.e. at least 60%) after two unprovoked seizures, occurring over the next 10 years; and 3) diagnosis of an epilepsy syndrome [1].

Epilepsy accounts for a high proportion of the global disease burden, with 13 million disability-adjusted life years (a summary measure of health loss defined by the sum of years of life lost for premature mortality and years lived with disability) [2]. Every year, more than five million new cases are diagnosed, and the number of people with epilepsy is expected to increase in future years [3]. This is attributed to the rising life expectancy worldwide and the increasing proportion of people surviving epilepsy-provoking insults, such as birth injury, head trauma, stroke, brain infection, and brain tumours.

The incidence rate of epilepsy is 61 per 100,000 person per year [3]. The incidence is higher in low and middle income countries than in high income countries. This can be explained by a greater exposure to perinatal risk factors, higher rates of central nervous system infections, and traumatic brain injuries. The prevalence of epilepsy is 7.6 per 1,000. It is also higher in low and middle income countries than in high income countries [3]. Incidence and prevalence are slightly higher in men than in women. The incidence of epilepsy is higher in young infants ($<$ 1 year old) and in the elderly. It has been estimated to be 86 per 100,000 during the first year of life, with subsequent decrease to 23-31 per 100,000 in the 30-59 age group, followed by increase to 180 per 100,000 in the age group above 85 [4]. In the last decades, the incidence of epilepsy has decreased in the younger age groups, probably due to improvements in perinatal care, better general health, and improved control of infectious diseases [2]. In contrast, the incidence has increased in the elderly, likely due to improved life expectancy with parallel increase of ageing-related epileptogenic conditions, such as stroke, tumours, and neurodegenerative disorders.

Seizures are classified according to their origin in the brain, the degree of awareness (knowledge of self and environment) and the level of body movement

during the seizure. Seizures are considered as focal if they originate within one hemisphere, usually in a group of neurons that are damaged or otherwise abnormal. They are regarded as generalised if seizure onset involves bilaterally distributed networks of the brain, with (quasi) instantaneous involvement of the entire brain [5].

Focal seizures occur when abnormal electrical activities originate from one cerebral hemisphere. If the patient is fully aware of the events occurring around them during the seizure it is categorised as a focal aware seizures. Most focal seizures induce a loss of awareness of some of the events during the seizure and are categorised as focal impaired awareness seizures. Examples of focal seizures include seizures arising from the temporal lobe, which are typically characterised by behavioural arrest and impaired awareness, often accompanied by oral and / or manual automatisms. Temporal lobe seizures start with sensory (auditory), emotional (fear), cognitive (déjà vu) or autonomic features (epigastric sensation, tachycardia, skin colour change) prior to onset of impaired awareness. Right after the seizure, patients typically show confusion. Patients with focal aware seizures may start with an aura experience, such as déjà vu, a strange taste or smell, sensation in the stomach, lip smacking, or and hand rubbing. Focal seizures can progress to invovle both hemispheres [6].

Generalised seizures can be categorised based on clinical manifestations, as motor or non-motor. The motor seizures include tonic-clonic, clonic, tonic, myoclonic, myoclonic-tonic-clonic, myoclonic-atonic, atonic, or epileptic spasms. Non-motor seizures are either typical or atypical absence seizures, or seizures with myoclonic activity or eyelid myoclonia. Most of the generalised seizures are associated with impairment of awareness, and therefore, the classification does not take awareness into account. Motor seizures are characterised by involuntary muscle movement that can manifest as jerking, tone stiffening or limb drop. Typical absence seizures (previously designated as 'petit mal'), are generalised seizures with abrupt onset and termination of altered awareness that commonly last about 10 to 20 seconds: the patient stares blankly and is not aware or responsive. Right after the absence, the person resumes normal activity. Often, they do not know that they had a seizure. Absence seizures are regarded as non-motor generalised seizures, although subtle clonic movements can occur at the level of the face or head, and oral and manual motor automatisms can be observed. Absence seizures are typically (but not exclusively) seen in a genetic or idiopathic generalised epilepsy syndrome known as childhood absence epilepsy. This epilepsy syndrome occurs in otherwise typically developing children, presenting with multiple daily absence seizures [7].

Seizures can occur at any moment without any notice. This weighs further on the burden of the disease. Prediction of seizure occurrence has been a hot research topic for the last 50 years in view of designing immediate, real-time anti-seizure

intervention [8], [9]. This research has fostered collaboration between multiple disciplines, namely neurology, electrophysiology, neuroscience, mathematics, physics, engineering, first-line health care, and patients and families. However, to date seizures cannot be predicted with sufficient accuracy and reliability to meaningfully improve treatment of patients with epilepsy [10].

The diagnosis of epilepsy is based on clinical history, neuroimaging and EEG. Relevant features include details of the events as they were experienced by the patient and witnesses, general context including past personal and family medical history, and specific circumstances under which the events occurred including possible triggers [7]. The clinical history is complemented by laboratory tests. These include blood and urine testing for evaluation of children [11], genetic testing [12], neuroimaging and most importantly EEG [13]. EEG can provide supporting evidence of epileptic activity, assist with classification of seizures and epilepsy conditions, provide prognostic information regarding seizure recurrence or following antiepileptic drug withdrawal, and plays a role in monitoring and treatment adaptation.

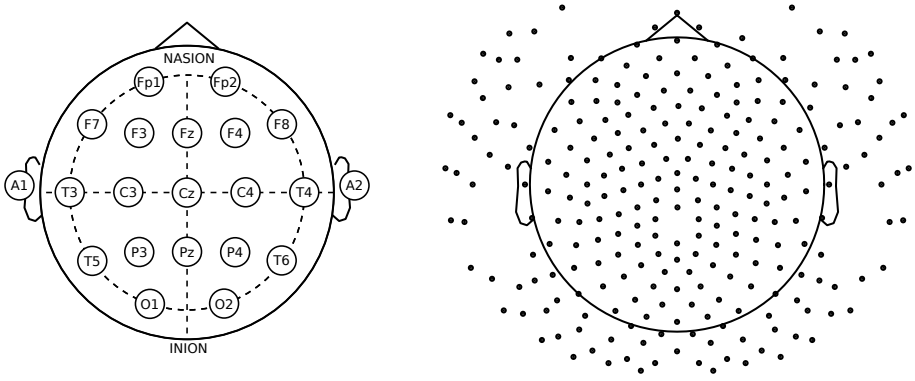
1.2 Electroencephalography

Scalp EEG is an electrophysiological measurement technique. It records the electrical field of the brain cortex non-invasively by placing a number of surface electrodes onto the scalp, typically about 20 in routine clinical practice. The brain electrical field reflects cerebral function involving the transmission of electrical signals by neurons (and possibly also glial cells) by ionic currents that flow through brain tissue. This activity is the result of both slow postsynaptic potentials (time scale 5 to 50 ms) and much shorter action potentials (1 ms). Postsynaptic potentials are the main contributor to EEG. Electrodes measure the summation of the activity of many neurons, implying that it is mostly synchronised activity of aligned neuron bundles that is being registered. Therefore, the EEG signal largely corresponds to the synchronised recruitment of large populations of neurons. Thus, relatively high amplitude EEG (in the range of hundreds of microvolts) mostly occurs during states of increased neuronal synchrony, such as drowsiness, sleep, and in some pathological conditions, such as epilepsy. Conversely, EEG attenuation is associated with desynchronisation of neurons.

High quality EEG recordings require low impedance between the reference and the measurement electrode. This impedance is due to the varying tissues and interfaces between the brain and the conductive electrode. Most of the total impedance is found in the layer of dead skin (stratum corneum) on the surface of the scalp. To record EEG, the skin is first cleaned with an abrasive gel to remove part of the stratum corneum. Then, electrodes are placed onto the scalp, with conductive gel to secure the electrode connection [14]. EEG electrodes are made of silver coated with a layer of silver chloride (Ag/AgCl), and these are typically stable, but other materials can be used, such as gold, platinum, tin, plastic with silver chloride inserts, wires, or epoxy coatings. These different steps result in minimising impedance below 10 kΩ [15]. Recording electrodes are connected to an amplifier, which converts small brain signals in the range of tens to hundreds of microvolts to a digital medium.

A clinical EEG hardware system typically uses about 20 channels, each corresponding to the electric field potential between two electrodes. This allows to cover cortical regions across the whole scalp. The international 10–20 system, first proposed in 1958, is used to place the ∼20 electrodes in a reproducible manner [16]. This system describes locations on the surface of the head relative to cranial landmarks. It places the first electrode at 10% of the distance between the nasion (i.e. the anatomic landmark at the junction between the forehead and the nose) and the inion (at the rear lower part of the skull) in the front to back direction and subsequent electrodes at 20% of the

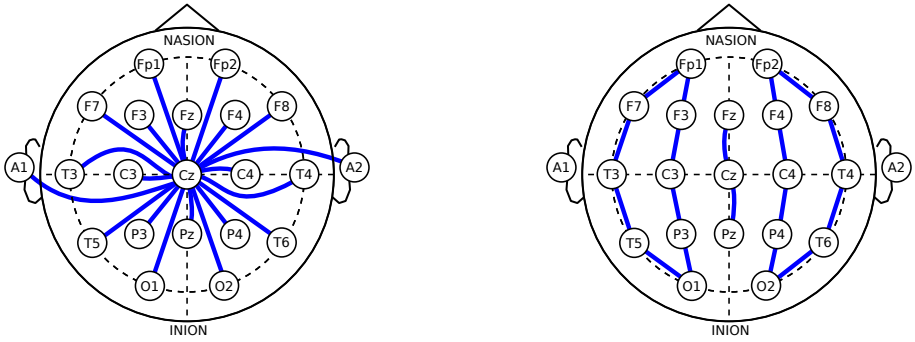
distance. In the right to left direction, the same procedure is used with the first electrode placed at 10% of the distance between the left and right mastoid in the right to left direction and subsequent electrodes at 20% of the distance. The international 10–20 system has been used almost universally and satisfactorily when the need for accurate electrode placement and high spatial resolution is limited. This system has been extended to higher density electrode settings, referred to as high density electroencephalograph (HD-EEG). Such extensions are sometimes used in the context of research and for the purpose of source localisation of EEG activity. The 10–10 and 10–5 systems have been designed, respectively, for at least 60 electrodes and at least 100 electrodes [17]. Even higher densities are used for inverse source imaging. This is also the case in the present work (257 electrodes). These different systems are shown in figure 1.1. Placement systems for more than 250 electrodes are provided by vendor-specific EEG head caps. In this work we use the EGI ™ Geodesic Sensor Net. This systems uses 256 channels with electrodes arranged along geodesic lines on the scalp (i.e. lines of shortest length along the surface of a sphere). We use this setup in chapter 4 to investigate the limits of miniaturisation of EEG.



a. 10-20 EEG b. EGI Geodesic Sensor Net

Figure 1.1: Placement of EEG electrodes on the scalp. a) 10–20 system of electrodes with 21 electrodes. b) high-density EEG system from EGI ™ Geodesic Sensor Net with 257 electrodes. *Electrodes displayed outside the scalp midline are located below the scalp midline.*

An EEG channel signal is obtained from the electric field potential between two electrodes, one being regarded as a measurement electrode and the other as a reference electrode. This allows recording and analysis EEG in different configurations, known as an EEG *montage*, depending on the choice of reference electrode. EEG can be recorded using a common reference montage where all electrodes are referenced to a single reference. This reference can be one particular electrode or the calculated average of all electrodes. If the electrodes are chosen appropriately, a localised EEG signal will be visible on the channel corresponding to the electrode overlying the cortical activity. EEG can alternatively be recorded using a bipolar montage. In such a montage, each channel is the potential difference within separate pairs of electrodes. Bipolar montages are more efficacious for recording localised EEG activity. These two different montage systems are shown in figure 1.2. These montage systems can either be physically connected in the EEG recording device or computed after the recording in a process referred to as re-rereferencing.



a. common reference (`Cz`)

b. bipolar montage

Figure 1.2: Commonly used EEG montage in a 10–20 EEG system. A blue line between two electrodes represents an EEG channel a) Common reference montage. All electrodes are referenced to electrode `Cz`. b) Bipolar montage. This bipolar montage is commonly named a 'double banana' montage.

The acquired EEG signal can be described in terms of spatial extent, frequency content, and morphological features. Signal oscillations are typically categorised into conventional frequency bands [18]:

- δ (< 3.5 Hz), e.g. prominent in the fronto-central regions during deep sleep.
- θ ($4 - 7.5$ Hz), e.g. prominent in fronto-central regions during drowsiness and early stages of sleep, or in frontal regions in children and young adults during heightened emotional states.
- α ($8 - 13$ Hz), e.g. the historically first described α rhythm that is characteristically present in the occipital head region in awake subjects when they close their eyes, and attenuated by eye opening and mental effort.
- β ($14 - 30$ Hz), which is the prevailing oscillation in typical recordings, particularly in frontal and central regions, with higher amplitude during drowsiness and early stages, and in people taking sedative medications such as barbiturates and benzodiazepines, which are also used as anti-seizure medications.
- γ (> 30 Hz), e.g. associated with sensory perception integrating information from different cortical areas.

In addition to background oscillatory activities, EEG episodic transients can be identified as isolated waveforms or complexes. Normal transients must be distinguished from pathological ones, particularly those signalling epilepsy. Identification of epileptic waveforms requires training and experience. Misinterpretation of non-epileptic transients as epileptic discharges can lead to overdiagnosis of epilepsy and unnecessary treatments that can induce adverse effects. There are two main markers of epilepsy in the EEG waveform, namely interictal epileptiform discharges (IED) and epileptic seizures. Criteria to identify IED have been established by the International Federation of Clinical Neurophysiology [19]. These criteria define an IED as:

1. Di- or tri-phasic waves with sharp or spiky morphology (i.e. pointed peak).
2. Different wave duration than the ongoing background activity: either shorter or longer.
3. Asymmetry of the waveform: a sharply rising ascending phase followed by a more slowly decaying descending phase, or inversely.

4. The transient is followed by an associated slow after-wave.
5. The background activity surrounding the IED is disrupted by the presence of the IED.
6. Distribution of the negative and positive potentials on the scalp suggests a source of the signal in the brain, corresponding to a radial, oblique, or tangential orientation of the source.

These six criteria are illustrated in figure 1.3. IED can appear as isolated spikes, sharp waves, polyspikes, spike and slow wave complexes, or brief potentially ictal rhythmic discharges (BIRDS). IED are the main diagnostic marker of epilepsy on the EEG as people with epilepsy have many IED between seizures, so that routine clinical EEG has a high chance of recording them. In a study conducted in 822 patients referred for EEG in support of the diagnosis of epilepsy, 36% of 30-minute long recordings showed at least one IED [20].

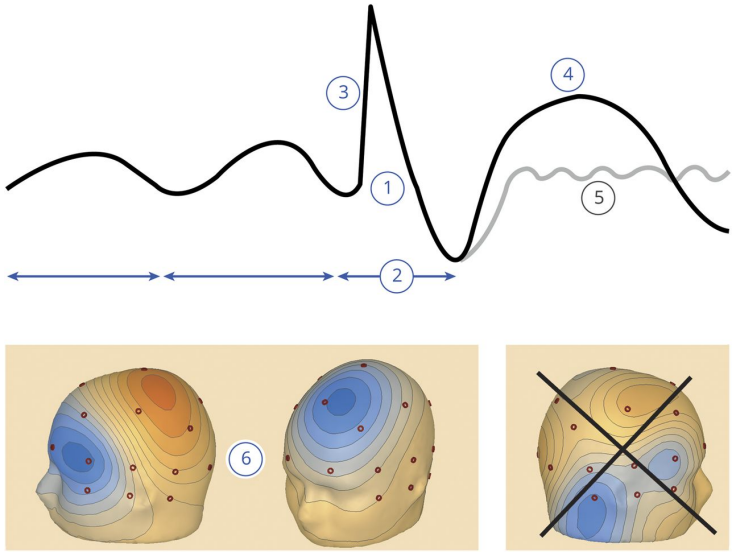


Figure 1.3: The six criteria of the International Federation of Clinical Neurophysiology for the definition of an interictal epileptiform discharge. The figure is reproduced from Kural et al. [19].

Ictal EEG patterns typically involve runs of rhythmic monomorphic discharge activity associated with clinical manifestations. The former are often more dramatic and more easily recognisable than IED, but seizures are classically of much rarer occurrence than IED, and their clinical features commonly induce major motion and muscle artefacts that jeopardise EEG interpretability. The patterns can be varied, and their recognition plays a role in diagnosis, sometimes supporting specific identification of a certain epileptic syndrome. In clinical practice, localisation of seizure onset (focal, generalised) is of great importance. This period is less prone to motion and muscle artefacts, and this localising information is often more relevant than that provided by IED [21]. There may also be recognisable immediate postictal changes, typically slowing of the background activity. Seizure patterns vary between epilepsy types and between patients. In idiopathic generalised epilepsy, generalised spikes and polyspikes are observed along with slow wave discharges at 3–5 Hz. The background cerebral activity is typically normal. In childhood absence epilepsy, bilateral synchronous 3 Hz discharges lasting 3–30 seconds are observed. The discharges are often slightly faster than 3 Hz at onset, and tend to slow down towards the end. The interictal EEG is normal, or may show runs of occipital rhythmic δ oscillation. In juvenile myoclonic epilepsy, the interictal and ictal EEG characteristic is brief bursts of polyspikes (sometimes a single spike). Variable asymmetry or lateralised emphasis of discharges is common, and interictal focal abnormalities are described in up to 40% of cases. Temporal lobe epilepsy associated with hippocampal sclerosis shows temporal interictal spikes and a characteristic rhythmic 5–7 Hz ictal discharges. In familial temporal lobe epilepsy, in contrast, focal IED is uncommon. [22]. An example of a generalised and focal epileptic seizure is shown in figure 1.4.

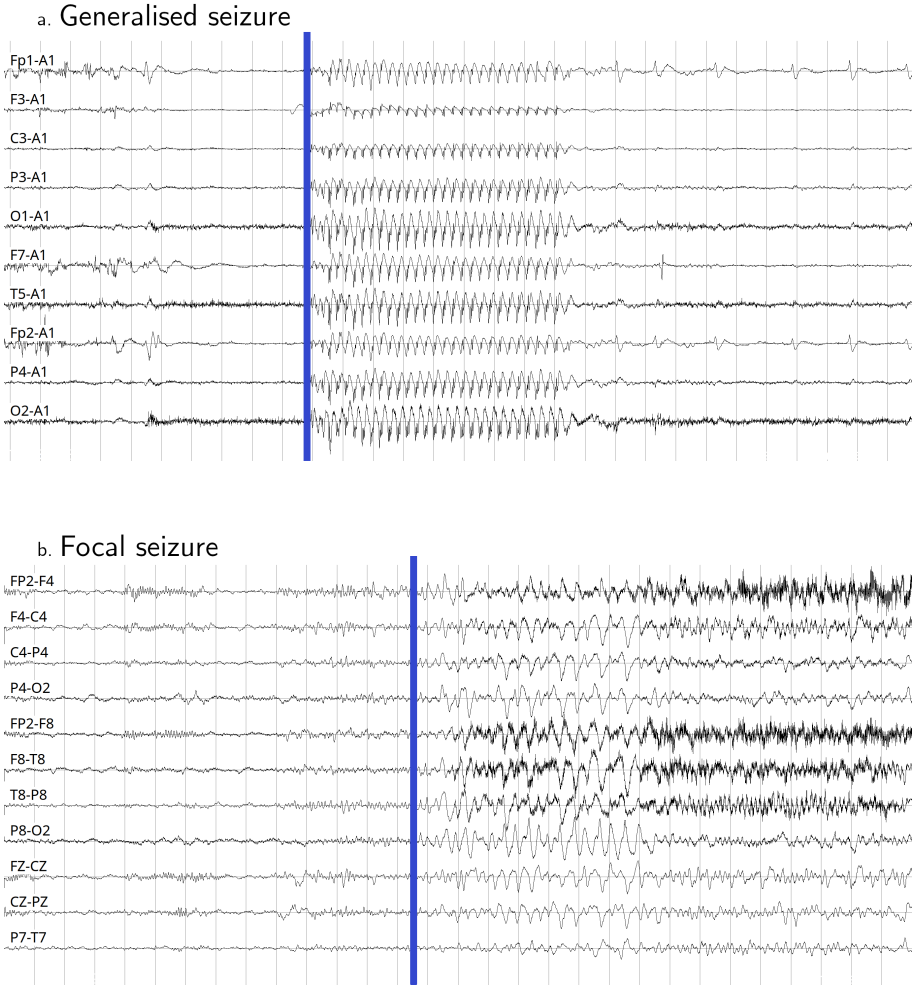


Figure 1.4: Examples of epileptic seizures. For each example 10 channels and 30 seconds of data are displayed. 500 μV separate two lines of data. Seizure onset is indicated by a blue line. a) A generalised seizure from a patient with childhood absence epilepsy. The seizure is characterised by generalised spike-wave discharges of approximately 3.5 Hz. The seizure lasts 8 seconds. The data were obtained from the study presented in chapter 2. b) A focal seizure with onset in the right parietal lobe. The seizure quickly spreads widely over the right hemisphere. The ictal EEG activity evolves from slow complexes (2.5 Hz) to fast (θ) activity. Several electrodes are contaminated by muscle artefacts during the ictal activity, likely corresponding to the motor manifestation of the seizure. The data were obtained from the public CHB-MIT Scalp EEG Database [23].

There is substantial variation in EEG recording practice in clinical settings [24]. Current requirement suggested by the ILAE for 'advanced' recording, i.e. standards that should be expected for tertiary epilepsy centres and university hospitals in resource-rich countries, include a digital device for multiple channel polygraphy, with video and activation procedures in addition to hyperventilation and intermittent photic stimulation, while stressing that all recordings should be individualised according to clinical information and questions. They also suggest that sleep EEG be recorded after sleep deprivation or pharmacological induction, and that prolonged (daytime or overnight) recordings are considered as indicated, ideally with video monitoring. The American Clinical Neurophysiology Society guidelines require a minimum of 20 minutes of artefact-free EEG recording for epilepsy diagnosis. However, longer recordings are more likely to yield a conclusive diagnosis. It was for example shown that recording durations of 40 minutes increased the diagnostic yield by 11% [25]. Long-term EEG monitoring is defined as recordings lasting more than 24 hours. It has been shown useful to classify epilepsy, leading to optimized treatment strategies with positive impact on the course of disease. In a 2021 study, Mann et al. showed long-term EEG monitoring altered the diagnosis of 37.4% of patients. They also reported it was well tolerated in 75% of patients and families [26].

1.3 Ambulatory EEG

While in-clinic EEG is an essential tool for epilepsy diagnosis, it has some limitations due to the short and episodic nature of the recordings. Some of these limitations can be addressed by using long-term ambulatory EEG. Long-term recordings provide more opportunities to distinguish between epileptic and non-epileptic events. Long-term EEG also increases the probability of recording epileptiform events when standard in-clinic EEG is non-diagnostic. Ambulatory EEG allows identification of seizure triggers that would be missed in a hospital setting. It also provides more realistic insights into the epilepsy condition than brief, episodic recordings in a highly standardised setting, therefore allowing decision-making regarding adjustment of anti-seizure medications [27], [28].

Ambulatory EEG monitoring comes with new challenges which are not present in in-clinic video-EEG. Ambulatory EEG typically does not include a video recording. This can make it challenging to distinguish between artefacts and epileptic activity. Given that subjects are being recorded in a free-moving environment, muscle and movement artefacts are a major challenge. Artefacts are also due to the difficulty in maintaining a low impedance. In the clinic, an EEG technician is responsible for checking the impedance of electrodes and re-applying gel when needed. This is more difficult to monitor and correct in

ambulatory EEG [29]. Ambulatory EEG recording devices also need to adhere to specific design requirements. They should encompass memory capacity that is large enough to record data over at least 24 hours. Devices should be light to allow free-range motion. The American Clinical Neurophysiology Society recently published minimum technical requirements for performing ambulatory EEG [27]. They recommend ambulatory EEG devices should record at a minimum sampling rate of 256 Hz. Electrode impedance should be less than 10 $k\Omega$ at the beginning of the recording and device weight should not exceed 500 grams [27].

Building such devices has been made possible by the advent of silicon integrated circuit chips and the subsequent miniaturisation of electronics. Early portable EEG devices focused on battery powered devices mimicking hospital-based systems. These devices, such as the Morpheus ™ EEG recorder from Micromed, allows monitoring of people with epilepsy in an ambulatory setup [30]. Initiatives around the world are transferring some of their standard practice of in-hospital video-EEG examinations to home EEG telemetry [31], [32]. While these devices make it possible to perform ambulatory EEG recordings, they remain bulky and inconvenient for unobtrusive use in everyday life. Advances in electronics also facilitates the development of cheaper consumer- and research-oriented devices. Devices such as the EMOTIV EPOCH allow the research community to build brain computer interface (BCI) [33] and gaming applications while devices such as the Muse headband can support consumer grade meditation applications [34]. More recently, companies such as Byteflies have developed miniaturised EEG devices that could pave the way for home monitoring of people with epilepsy in their everyday life [35]. These different devices are illustrated in figure 1.5. In addition to these commercially available products, research has focused on different aspects of the miniaturisation of EEG devices. Active research is also taking place on electrode technology. It aims at improving ease of use to allow users to self-apply the electrodes. These electrodes should also allow for long-term ($>$ 24 hour) monitoring with minimal user interaction. This may require the development of so-called dry electrodes that do not make use of water-based solutions [14]. Manufacturing techniques have been developed to produce arrays of miniature needles [36], [37]. These microneedles are several hundreds of μm long and have a diameter of less than 100 μm. The working principle is that they pierce the stratum corneum of skin, thereby greatly reducing impedance compared to flat dry electrodes. In parallel, amplifier technology is continuously being developed in order to achieve lower power consumption while maintaining high gain and high input impedance for use with new dry electrodes [38]. These modern integrated circuits (ICs) are then packaged into wearable form factors. These can take the form of behind the ear EEG [39] or in-ear EEG where electrodes are placed within the ear attached to an ear bud [40].

Other efforts have designed electronics integrated into the frame of eyeglasses for discrete everyday use [41]; still others have developed headbands [42]. Another promising approach consists of implanting an array of electrodes subcutaneously in a minimally invasive surgical procedure. This can allow continuous EEG recording during extended periods of time, up to many months [9], [43].

Figure 1.5: Examples of different portable EEG devices. A) Neurosoft home-video telemetry setup for home-video EEG monitoring. B) EMOTIV EPOCH amplifier for use as a research device in mobile environments. C) Byteflies Sensor Dot for EEG recording in people with epilepsy in their everyday life. *(Images are reproduced with the authorisation of the manufacturers. Image rights remain the property of the device manufacturer.)*

Traditional ambulatory EEG rely on a centralised system where all electrodes are connected to a central recorder. This requires wires between the electrodes and central recording unit which are the cause of artefacts and hinder wearability. An alternative platform has been proposed to improve wearability [44]. This new platform consists of a collection of wireless miniature EEG sensor units operating as a sensor network. Each EEG unit should incorporate electrodes, an amplifier, a wireless radio and a processing unit in single package with a small scalp area footprint. It is illustrated in figure 1.6. Electronics designed for a wireless EEG sensor network have been developed by several groups [45], [46] and algorithms making use of this type of platform have been proposed for BCI applications [47]. In this thesis we rely on algorithms that can be expressed as a generalised eigenvalue decomposition (GEVD). Musluoglu et al. proposed a framework for solving spatial filtering optimization problems in a distributed fashion [48]. The previous sections of this introduction showed most components required for a practical realisation of a wireless EEG sensor network are already available or in active development. In this thesis, we investigate how a wireless EEG sensor network can be used in the context of home monitoring of people with epilepsy. For this, we develop signal processing methods for the analysis of EEG signals recorded in a wireless EEG sensor network. We also develop algorithms to design such a network in the context of home monitoring of people with epilepsy. Next to the EEG amplifier technology and signal processing algorithms, electrodes for home monitoring of EEG are also required. These are still in active development and not yet validated for long term monitoring.

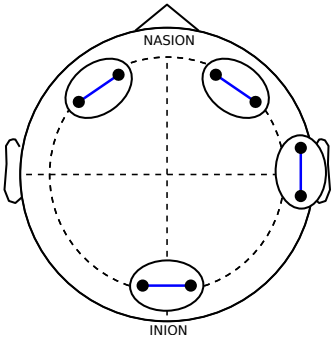


Figure 1.6: Example of a wireless network of galvanically isolated EEG units. This example network consists of four mini-EEG units. Each unit incorporates a bipolar channel with two electrodes, an amplifier, a wireless radio and a processing unit in a small package.

1.4 Signal Processing of the Electroencephalogram

In this dissertation, we develop new signal processing methods for the analysis of EEG signals recorded in people with epilepsy. These algorithms produce human-interpretable output from the large volume of data recorded by multiple channels of EEG sampled at more than 200 Hz. The aim of this introductory section is to present common signal processing tools for EEG analysis, as well as the background to understand the main chapters in this thesis.

The first step for EEG analysis consists of pre-processing the signal in order to format the raw EEG data for analysis independently from individual recording settings. This requires making sure that the data are appropriately sampled, referenced and filtered, as well as eliminating artefacts that would hamper signal analysis.

First the EEG must be acquired at a sample rate that captures all the relevant information in the EEG and ensures the recorded signal does not suffer from aliasing effects. To achieve this, the sampling frequency (fs) should be twice as high as the frequency of the highest signal, according to the Nyquist–Shannon sampling theorem, as any signal above this frequency will be aliased. To ensure proper sampling of the signal, an EEG amplifier should implement a hardware low-pass filter prior to sampling. The signal is then filtered to the frequency band of interest. This is typically done using a combination of high-pass, low-pass, and notch filters. The high-pass filter removes the direct current (DC) offset and baseline wander as these are caused by the electronic and slow impedance changes that are unrelated to neural activity. To remove the DC offset and baseline wander, a zero-phase infinite impulse response (IIR) high-pass filter with a cut-off frequency of 0.5 Hz is typically used. The low-pass filter uses a similar filter design with a cut-off dependent on the application. They can be set to a cut-off frequency of 35 Hz in the presence of prominent muscle artefacts or 70 Hz in more conservative settings. A notch filter is used to remove the power-line interference. The EEG can then be resampled to a lower (or standardised) sampling frequency depending on the required application. For example, in chapter 4 the EEG was acquired at 1 kHz, however the analysis did not require such a high sampling rate such that the signal was resampled to 100 Hz.

Artefacts in the signal should be dealt with in order to allow the signal processing pipeline to achieve its task. Artefacts, such as those induced by disconnected electrodes, can reach values that are several orders of magnitude larger than physiological neural activity recorded by the EEG. Artefacts can have a physiological origin, e.g. blinking, eye movement, muscle contractions, sweat, cardiac electrical activity, pulse, or body movement. Artefacts can also have

a non-physiological origin in the power-line interference, changes in electrode impedance, tapping or movement of electrode wires or electrodes, and electrode disconnections. Artefacts are a challenge for the analysis of EEG as they occult the underlying neural activity. They are often challenging to remove without distorting the signal reflecting neural activity. Several strategies are used to deal with artefacts. They include manual or automatic rejection of epochs or channels severely contaminated by artefacts. The latter is the procedure we use in this thesis when selecting the beginning and end of a recording as well as rejecting epochs with abnormally high recorded values. Specific artefact rejection algorithm can be included in the pre-processing pipeline. Common examples include techniques based on wavelet decomposition, independent component analysis, and Wiener filters [49]. However, one must accept that some artefacts remain in the signal and they are part of the challenges for a signal processing pipeline. Controlled environments, such as those organised in the hospital, tend to induce less artefacts than recording conditions in environment in which people can move freely.

The task of detecting epileptiform events is central to this thesis. It can be formulated as a classification problem that aims at providing binary labels to data epochs: "epileptiform event" or "non-epileptiform event". This task is often approached as a classical machine learning problem where a bag of features that should allow discrimination between the two states are extracted from the signal and fed to a classifier that determines a decision boundary based on the value of the feature set [50]. Many features have been proposed in the literature as relevant for the detection of epileptiform events [51]. They correspond both to linear and non-linear descriptors of the signal [52], [53]. They relate either to the time domain, the frequency domain, to the spatial domain (across channels), or a combination of the former dimensions. Since the advent of neural networks, complex neural network architectures have been developed to operate directly on the raw signal [54]. Some of these have been shown to outperform classical classifiers [55]. In addition to feature extraction or classifiers operating on raw data, filters can be designed to attenuate activity that is unrelated to epileptiform activity. The output of these filters can then be used to classify epochs. This strategy is attractive in that it allows for a tight control of the computational complexity in function of the number of filter parameters. It also allows for early fusion of the multi-channel EEG as raw signals are fused together to produce a single output rather than extracting per-channel features. This allows to exploit the spatio-temporal statistics of the EEG signal. This is the strategy used in this thesis. It is explained in detail in chapter 2.

In this thesis we make the assumption the EEG signal is stationary. This means the first order statistics of the signal do not vary over time. This assumption is often found in signal processing algorithms and is a prerequisite for many mathematical problems. However, the stationarity assumption is easily violated in practice. Several causes can result in non-stationarity of the EEG signal statistics. Among them, changes in electrode impedance lead to degradation of the signal quality and changes in the signal statistics. Changes in brain states could also influence the signal statistics, but should only have a minimal impact on the filter performance if training data was presented with these different states. Everyday life activities can also lead to changes in the statistics of the signal. Repetitive artefacts such as those found while walking or vibration associated with motorised transportation can lead to non-stationarity of the EEG. The recording conditions in this thesis should guarantee a relative stationarity of the EEG signal. However, we caution the reader to keep in mind stationarity might not be guaranteed in some long-term EEG recordings.

1.5 Research objectives and thesis overview

1.5.1 Research objectives

In the previous sections of the introduction we highlighted the health burden on people with epilepsy. We explained how EEG is the main diagnostic tool for epilepsy. We detailed the current clinical standard for EEG measurements in the hospital and enumerated current efforts to develop wearable EEG for home monitoring of people with epilepsy. However, many current EEG wearable devices still suffer from usability issues. One of the major limitations relates to the trade-off between wearability and number of electrodes. More electrodes lead to larger electronics and more wires, which make the platform less user-friendly, and induce more wire-related artefacts. Conversely, the less electrodes the lower the spatial resolution and the less tasks the system can be used for.

We propose a different platform to offer solutions to some of the current problems associated with wearable EEG devices. This new platform will consist of a collection of wireless miniature EEG sensor units operating as a sensor network. Each EEG unit should incorporate electrodes, an amplifier, a wireless radio and a processing unit in single package with a small scalp area footprint. Each unit would record a single local bipolar EEG channel with the units placed anywhere on the scalp. The new modular topology eliminates the need for wires and will allow for a flexible, discreet, miniature, wearable system with as many EEG sensors as necessary for a particular application or patient. This absence of wires removes all physical connections between EEG units resulting in galvanically

isolated EEG units. This wireless network of galvanically isolated EEG units should also suppress artefacts resulting from wire movement and interference [56]. This novel modular approach would facilitate a flexible miniaturised design with multi-channel recordings at various positions on the scalp, but it will also require new signal processing tools to analyse the EEG signals. To limit the energy consumption of our system, we will design specific algorithms for embedded applications. The new opportunities offered by the system can hopefully be used for automated detection of seizures and interictal epileptic discharges from multi-channel EEG in daily life, which is important for diagnosis and follow-up of people with epilepsy.

The concrete objectives of this thesis are :

1) Design algorithms for real-time automatic analysis of epileptiform activity in resource-constrained (wearable) EEG platforms. These algorithms should automatically detect relevant physiological markers of epilepsy on EEG such as epileptic seizures and interictal epileptiform discharges. The algorithms should be designed to operate on limited hardware resources to allow real-time applications such as providing an alarm to a caregiver in case of a seizure. They must take into account the limitations on computational resources, battery lifetime, wireless bandwidth, random access memory (RAM), and storage, as well as the limitations resulting from the miniaturisation of the EEG setup. This entails short inter-electrode distance resulting in low signal-to-noise (SNR) as well as a limited number of EEG channels. These algorithms assume stationarity of the EEG signal. They should be designed to operate in free ranging environments where artefacts are common occurrence.

2) Design and benchmark strategies for optimal EEG sensor selection. The number of available EEG sensor units generates a trade-off between performance and wearability such that the number of sensors should be minimised while ensuring sufficient performance for the task at hand (e.g. automatic epileptic seizure detection). The EEG channel selection should be optimised for epileptiform activity detection while optimally reducing noise. It should be computationally tractable such that the channel selection on a large number of input variables can be computed in reasonable time to allow many simulation scenarios such as those explored in chapter 4.

3) Provide guidelines for the design strategies of miniaturised EEG sensor networks. Simulations of networks of miniaturised EEG networks should provide limits of miniaturisation of such networks, both in terms of inter-electrode distance (2–8 cm) and number of units (1–10 units) in a network.

1.5.2 Chapter overview

This thesis consists of three main chapters. A schematic overview of the thesis is provided in figure 1.7. The first chapter introduces a computationally efficient algorithm for absence seizure detection in the context of wearable EEG. This algorithm is based on a GEVD framework that serves as a basis for the rest of the work in this thesis. The results of chapter 2 will reveal the important issue of appropriate selection of EEG channels or electrodes. It leads on to the development of a channel selection method for GEVD problems, which is presented in chapter 3. The methods developed in chapters 2 and 3 are used to investigate the limits of miniaturisation of EEG for monitoring of epilepsy, which is addressed in chapter 3. These three chapters presenting novel original research are followed by conclusions and future research directions (chapter 5).

Chapters 2 and 3 are based (with minor modifications to optimise fluidity of the presentation of the dissertation) on peer-reviewed, published papers. All papers are first-authored by myself, in close collaboration with supervisors and research collaborators. A more detailed overview of the different chapters can be found below.

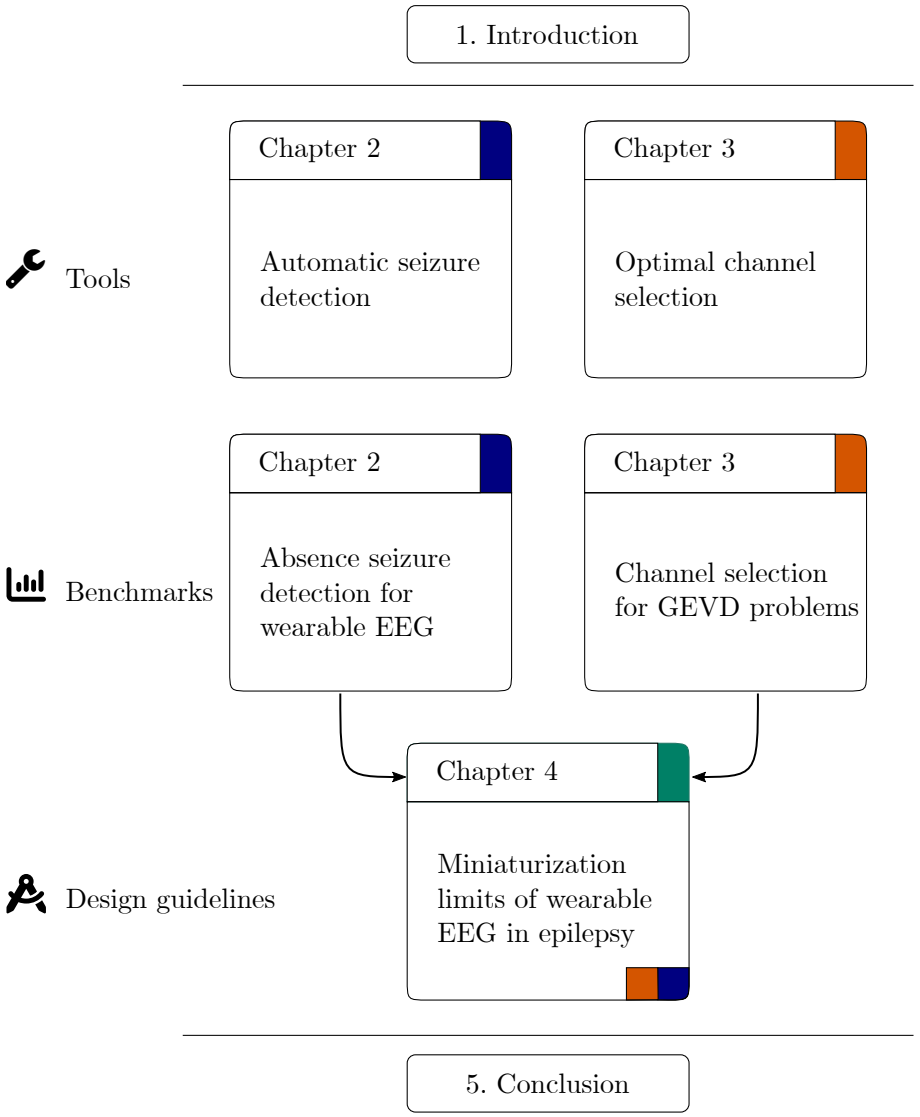


Figure 1.7: Schematic representation of the thesis text. The thesis is split into three main chapters. Chapter 2 develops an algorithm for automatic detection of absence seizure detection. The algorithm is based on the GEVD framework. This is the basis for the channel selection and benchmark of methods in chapter 3. The insights in chapter 2 and 3 are used in chapter 4 to analyse the effects of miniaturisation of EEG.

Chapter 2: Computationally-Efficient Algorithm for Real-Time Absence Seizure Detection in Wearable Electroencephalography describes the design of a novel computationally-efficient algorithm for absence seizure detection. The algorithm is specifically designed to run on microcontrollers as would be found in miniature, wearable EEG devices. The algorithm is benchmarked against other algorithms for absence seizure detection in the context of wearable EEG. The algorithm is based on a GEVD framework which is used throughout this thesis. The chapter highlights the importance of channel selection which is explored in more detail in chapter 3.

Chapter 3: Grouped Variable Selection for Generalised Eigenvalue Problems presents a new variable selection method for GEVD problems that can for example be used for channel selection in EEG. The chapter also compares different algorithms for variable context in the GEVD problems. This chapter is used to select the best channel selection method to explore the limits of miniaturization presented in chapter 4.

Chapter 4: Sensor selection and miniaturisation limits for detection of interictal epileptiform discharges with wearable EEG explores the limits of miniaturization of wearable EEG devices, both in terms of inter-electrode distance (2–8 cm) and number of units (1–10 units) in a network, by making use of the methods in chapters 2 and 3. The results from the chapter give design guidelines for miniature EEG devices.

CHAPTER 2

Computationally-Efficient Algorithm for Real-Time Absence Seizure Detection in Wearable Electroencephalography

This chapter is based on J. Dan, B. Vandendriessche, W. Van Paesschen, D. Weckhuysen, and A. Bertrand, (2020). "Computationally-Efficient Algorithm for Real-Time Absence Seizure Detection in Wearable Electroencephalography", *International journal of neural systems*, vol. 30, no. 11, 2020. A Graphical User Interface to improve review of EEG recordings from patients with absence seizures is added as an appendix.

Computationally-Efficient Algorithm for Real-Time Absence Seizure Detection in Wearable Electroencephalography

Abstract | Advances in EEG equipment now allow monitoring of people with epilepsy in their daily-life environment. The large volumes of data that can be collected from long-term out-of-clinic monitoring require novel algorithms to process the recordings on board of the device to identify and log or transmit only relevant data epochs. Existing seizure-detection algorithms are generally designed for post-processing purposes, so that memory and computing power are rarely considered as constraints. We propose a novel multi-channel EEG signal processing method for automated absence seizure detection which is specifically designed to run on a microcontroller with minimal memory and processing power. It is based on a linear multi-channel filter that is precomputed offline in a data-driven fashion based on the spatial-temporal signature of the seizure and peak interference statistics. At run-time, the algorithm requires only standard linear filtering operations, which are cheap and efficient to compute, in particular on microcontrollers with a multiply-accumulate unit. For validation, a dataset of eight patients with juvenile absence epilepsy was collected. Patients were equipped with a 20-channel mobile EEG unit and discharged for a day-long recording. The algorithm achieves a median of 0.5 false detections per day at 95% sensitivity. We compare our algorithm with state-of-the-art absence seizure detection algorithms and conclude it performs on par with these at a much lower computational cost.

2.1 Introduction

Epilepsy is one of the most common severely disabling brain conditions, affecting over 46 million people worldwide [57]. The clinical utility of monitoring epileptic seizures has been amply documented [58]. For example, it is used for optimizing antiepileptic therapy and for the development of new drugs. Until recently, long-term monitoring of epilepsy relied mostly on patient-reported outcomes and occasional short duration EEG recording in hospital-based epilepsy monitoring units to document epileptic discharges. Such in-hospital recordings rarely contain actual electroclinical seizures as these are relatively unlikely to occur during the very short recording sessions [59]. For a small proportion of people with epilepsy, typically those selected for surgical treatment, this can eventually be complemented by additional (one to several day-long) video-EEG recordings in specially equipped hospital units [58]. Epilepsy monitoring units confine recordings within a hospital room, imposing restriction on patients' movement and activity. These hospital recordings also incur significant expenses [60].

Epilepsy is classified in different syndromes and seizure types. Typical absence seizures are generalised seizures that occur in people with juvenile absence epilepsy and childhood absence epilepsy syndromes. These seizures are characterized by sudden impairment of consciousness lasting a few seconds [61]. The seizures are associated with a bilateral, synchronous, and symmetrical EEG discharge of 3 Hz spike and wave complexes [62]. Given the lack of distinctive movement or other clinical manifestations, it is difficult for people with absence seizures or their caretakers to accurately detect and count seizures, and report them to the clinician. An ambulatory EEG study in people with absence seizures found that only 6% of seizures lasting more than three seconds were reported [63]. Due to this severe under-reporting, a wearable EEG device that automatically detects and logs absence seizures in daily life situations would be a valuable asset for the people with epilepsy and the clinician.

The emergence of miniature EEG devices that can record EEG outside a hospital or lab environment enables ambulatory measurements in a real-life setting [39], [64]–[67]. This allows long-term recordings, which have a much higher chance of capturing seizures [68]. In addition, seizures resulting from epileptic triggers that would not occur in the hospital can be recorded, sleep patterns are not disturbed, and the neurologist has access to a more representative recording of the patient's seizures. In order for these new EEG devices to be accepted and worn by people with epilepsy outside the hospital for prolonged periods of time, the devices need to be designed to be comfortable, miniature, wearable and discreet. Such EEG devices could be used to monitor people with epilepsy over long periods of time, generating large amounts of data.

In turn, this would require automated data analysis to extract relevant biomarkers from large datasets as the amount of manual annotation work would not be manageable realistically by an epileptologist. Furthermore, the recorded EEG data have to be processed on the device itself in order to minimize the amount of data that have to be logged within the device or transmitted to the cloud. This would ensure a sufficiently long battery lifetime.

Many seizure detection algorithms have been proposed for use in the clinic [69], [70]. Most of these are not built for a specific seizure type and are operated offline on data collected during routine clinical observation in epilepsy monitoring units. Only a few algorithms have been designed specifically for people with absence seizures [71]–[75]. In even fewer of these studies, algorithms were designed to run on wearable systems. In contrast to algorithms developed to run in a hospital without strict computational constraints (as they can rely on powerful servers connected to the hospital network) [70], algorithms for wearable systems must meet with strict storage, computing memory, and computing power constraints. The challenge of developing energy efficient seizure detection algorithms has also been highlighted by teams working on long-term intracranial EEG for use in implanted devices [76]–[78].

In this chapter we propose a fully automated seizure detection algorithm specifically designed to run on a microcontroller. The algorithm is based on a linear filter that is designed in a data-driven fashion to maximally amplify absence seizure signals while optimally attenuating peak interference. This framework is inspired by work from Wouters et al. in spike sorting where a template matching algorithm is used [79]. Wouters et al. propose to optimize a data-driven filter in terms of separation of target signal and peak interferences, thereby maximizing the signal-to-peak interference ratio. This is an adaptation of classical data-driven filters that optimize signal-to-noise (SNR) ratio. It is then applied as a linear filter-and-sum operation with finite impulse response (FIR) filters. FIR filtering is a standard signal processing operation that can be efficiently implemented in low-power hardware, in particular when a dedicated multiply-accumulate (MAC) unit is available.

The chapter is organized as follows. The data collection and the data-driven filter design are presented in section 2.2. The results of the proposed algorithm are reported and compared to other state-of-the-art absence seizure detection algorithms in section 2.3. The results are discussed in section 2.4 and conclusions are drawn in section 2.5.

2.2 Materials & methods

2.2.1 Data collection and annotation

The main dataset used in this study consisted of recordings performed in out-of-clinic environments. An additional validation dataset was assembled from retrospective recordings obtained in a hospital setting to validate the generalizability of the proposed algorithm to new independent data sets with additional patients and different recording equipment.

For the main dataset, EEG signals were obtained from eight patients with refractory juvenile absence epilepsy recruited at the epilepsy reference center of the UZ Leuven university hospital (Belgium) (Table 2.1). Inclusion criteria were age of 18 or above, and diagnosis of refractory juvenile absence epilepsy ascertained by an expert neurologist. Exclusion criteria for data analysis were the absence of recorded absence seizures. Participants were equipped at the hospital with a 20 channel Medatec BrainWalker3 (Braine-le-Château, Belgium) portable EEG amplifier (10-20 system) sampling at 200 Hz and then discharged for ambulatory continuous EEG recording until the next day. During the recording, they were allowed to proceed with their daily life activities (the only restriction was not washing their hair).

Patient	Sex	Age	Seizure types	Recorded seizures	Seizure duration	Recording duration
1	F	49	A, GTCS, M	0	NA	05:53:30
2	F	25	A, GTCS	41	26	22:43:47
3	M	18	A, GTCS	8	10	21:45:00
4	M	24	A	0	NA	NA
5	F	42	A, GTCS	5	8	21:18:20
6	F	22	A, GTCS	5	4	21:55:33
7	F	20	A, GTCS	24	13	21:43:40
8	F	24	A, GTCS, M	18	9	21:50:40

Table 2.1: Patients with juvenile absence epilepsy recruited for 24 hours out-of-clinic EEG monitoring. Sex: F (female), M (male); Age in years; Seizure types: A (absence seizure), GTCS (generalised tonic-clonic seizure), M (myoclonic seizure); Median seizure duration in seconds

For the independent validation dataset, EEG recordings from 17 patients with refractory juvenile absence epilepsy who visited the epilepsy monitoring unit during the period 2016-2019 were aggregated (Table 2.2). The patients were recorded within the video-EEG unit for a day-long monitoring using a 20 channel BrainRT Brainbox 1042 EEG recorder (Kontich, Belgium) sampling at 250 Hz. Throughout those recordings, the patients stayed in the video-EEG monitoring room. These recordings are of high quality and contain little technical or movement related artifacts. The first eight patients from table 2.2 are the same as the patients in table 2.1.

Patient	Sex	Age	Seizure types	Recorded seizures	Recording duration
1	F	49	A, GTCS, M	5	21:59:25
2	F	25	A, GTCS	0	20:42:50
3	M	18	A, GTCS	0	20:16:50
4	M	24	A	5	20:38:04
5	F	43	A, GTCS	0	14:01:40
6	F	22	A, GTCS	0	20:56:13
7	F	20	GTCS	0	19:02:21
8	F	24	A, GTCS, M	0	20:12:47
9	M	28	A, GTCS, M	20	19:54:40
10	F	30	A, GTCS	2	21:48:35
11	F	41	A, GTCS	10	19:44:12
12	M	24	A	0	20:34:56
13	F	32	A	0	21:22:48
14	F	47	A, GTCS	88	21:16:33
15	F	49	A, GTCS	6	19:40:05
16	F	48	A	40	19:54:04
17	F	33	A, GTCS	43	21:54:56

Table 2.2: Patients with absence epilepsy recruited for 24 hours in-hospital EEG monitoring. Sex: F (female), M (male); Age in years; Seizure types: A (absence seizure), GTCS (generalised tonic-clonic seizure), M (myoclonic seizure)

Absence seizures, defined as electroencephalographic generalised spike and wave discharges, were annotated by an expert epileptologist from UZ Leuven. It is noted that epileptiform graphoelements with a duration of less than three seconds are generally associated with interictal activity [80]. It depends on the targeted use case whether such shorter events should be flagged by a detection algorithm or not, and so whether their detection should be treated as true or false positives. In this work, this choice was avoided by excluding all epileptiform segments with a duration shorter than three seconds from the analysis.

The study was approved by the local ethics committee and written informed consent was obtained from all participants.

2.2.2 Seizure detection algorithm

Our proposed seizure detection algorithm is a pre-trained spatio-temporal filter, implemented as a filter-and-sum pipeline; followed by thresholding on the power of the single-channel output signal. This approach is illustrated in Figure 2.1.

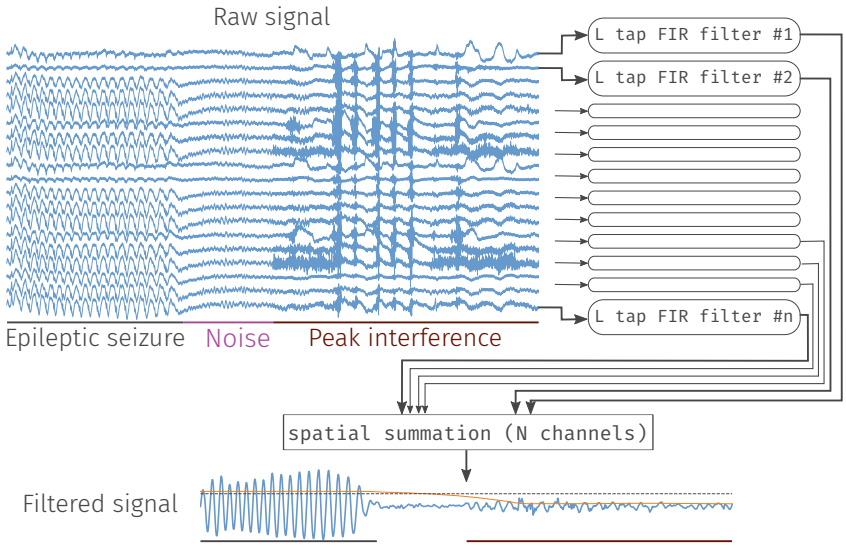


Figure 2.1: Illustration of the seizure detection algorithm. Bandpass filtered multichannel EEG signal (top left, blue) is fed to a pre-trained (see Fig. 2.2) spatio-temporal filter (right) implemented as a filter-and-sum pipeline where each FIR filter has L filter coefficients ('taps'). This results in a single-channel output (bottom, blue). The time-varying standard deviation (root mean square (RMS) value) of this filter output signal is computed (bottom, orange). All samples above a set threshold (bottom, black) are labeled as seizures.

The training of the filter coefficients is done in a two-step approach. First, peak interferences are detected by training and applying a purely spatial filter that maximizes the seizure power to non-seizure power ratio. The 40 minutes of highest power of the single-channel output (excluding seizure segments) are labeled as peak interference. Then, a second spatial-temporal data-driven filter that maximizes the seizure to peak interference ratio (SPIR) is trained. This filter can be adapted in terms of number of channels and number of time lags. A detailed explanation of the algorithm is given in the following subsections.

Data preprocessing

The data were high-pass and low-pass filtered with a fourth-order Butterworth filter with a cut-off frequency of 0.5 Hz and 25 Hz, respectively. The data were then downsampled to 50 Hz and re-referenced to a longitudinal bipolar montage [81] resulting in 18 channels. Segments with a RMS amplitude of more than 400 μV over a window of 100 ms were considered as non-electroencephalographic (i.e. not of brain origin) and excluded from the analysis along with 1.5 seconds of data before and after each excluded segment of 100 ms.

Data-driven filter design

The EEG signal in channel k is modeled as

$$y_k(t) = s_k(t) + n_k(t) + i_k(t)$$

where $s_k(t)$ corresponds to seizure-related EEG activity, $n_k(t)$ corresponds to background EEG or noise components that are not related to the seizure [1], and $i_k(t)$ corresponds to peak interference, i.e. artifacts that occur sparsely in time with a high peak amplitude, e.g. due to chewing or eye blinking. Peak interferers are typically the main source of false positives in a threshold-based seizure detection algorithm, whereas the noise floor denoted by $n_k(t)$ generally does not trigger a detection. Note that $s_k(t)$ and $i_k(t)$ are sparse processes, i.e. the signals are zero most of the time, whereas the noise floor signal $n_k(t)$ is continuously active. The raw data in figure 2.1 show periods where seizure activity, noise, and peak interference (chewing artifact) are respectively dominant.

Let $\mathbf{y}(t) \in \mathbb{R}^N$ denote a N-dimensional vector containing the sample at time t collected at N EEG channels, i.e. $\mathbf{y}(t) = [y_1(t) \ \dots \ y_N]^T$, where T denotes the transpose operator. The vectors $\mathbf{s}(t)$, $\mathbf{n}(t)$, and $\mathbf{i}(t)$ are defined similarly, such

[1] In the remaining of the chapter, we will refer to $n_k(t)$ as the 'noise' component. The term noise should be read in the signal processing terminology as non-target signal components, i.e., signal dynamics that are not related to a seizure.

that $\mathbf{y}(t) = \mathbf{s}(t) + \mathbf{n}(t) + \mathbf{i}(t)$. The N EEG channels are then linearly combined into a single-channel output signal $o(t)$.

$$o(t) = \mathbf{w}^T \mathbf{y}(t) \tag{2.1}$$

where $\mathbf{w} \in \mathbb{R}^N$ contains the combination weights. From a signal processing point of view, $\mathbf{w}$ acts as a spatial filter as it linearly combines different EEG channels at different positions on the scalp. Our goal is to find the optimal set of filter coefficients such that the filter output signal $o(t)$ has a maximal amplitude if a seizure is present, while suppressing the noise floor and peak interferers as much as possible. In other words, the filter $\mathbf{w}$ is optimized in a data-driven fashion to maximize the SPIR of $o(t)$ over a training set, i.e. solving

$$\max_{\mathbf{w}} \frac{E_s\{(\mathbf{w}^T\mathbf{s}(t))^2\}}{E_i\{(\mathbf{w}^T(\mathbf{i}(t) + \mathbf{n}(t)))^2\}} \tag{2.2}$$

where $E_s\{.\}$ denotes the expected value operator evaluated over epochs in which a seizure is present, and $E_i\{.\}$ denotes the expected value operator evaluated over signal segments during which a peak-interferer is active.

Equation 2.2 is equal to:

$$\max_{\mathbf{w}} \frac{\mathbf{w}^T\mathbf{R_s}\mathbf{w}}{\mathbf{w}^T\mathbf{R_{i+n}}\mathbf{w}} \tag{2.3}$$

where $\mathbf{R_s} = E_s\{\mathbf{s}(t)\mathbf{s}(t)^T\}$ is the seizure covariance matrix and $\mathbf{R_{i+n}} = E_i\{(\mathbf{i}(t)+\mathbf{n}(t))(\mathbf{i}(t)+\mathbf{n}(t))^T\}$, is the peak-interference-plus-noise covariance matrix. In subsection 2.2.2, we will explain how these covariance matrices are estimated from the data. It can be shown[82] that the solution of the maximization problem defined in equation 2.3 is the eigenvector corresponding to the largest eigenvalue of the matrix $\mathbf{R_{i+n}^{-1}}\mathbf{R_s}$.

The filter described above is a purely spatial filter. It can be expanded to a causal spatio-temporal filter by creating a buffer of L samples for each channel and stacking all buffered (time-lagged) samples in a single vector $\tilde{\mathbf{y}}(t) = col\{\tilde{\mathbf{y}}_\mathbf{1}(t), \ldots, \tilde{\mathbf{y}}_\mathbf{N}(t)\}$ where $\tilde{\mathbf{y}}_\mathbf{k}(t) = [y_k(t), y_k(t-1), \ldots, y_k(t-L+1)]^T$ and $col\{.\}$ denotes a columnwise stacking. The output signal $o(t)$ is given by equation 2.1 where $\mathbf{w}$ is replaced by $\tilde{\mathbf{w}} \in \mathbb{R}^{LN}$ and $\mathbf{y}(t)$ is replaced by $\tilde{\mathbf{y}}(t) \in \mathbb{R}^{LN}$. This then corresponds to a filter-and-sum operation as depicted in Fig. 2.1, where each channel is filtered with a channel-specific L-taps FIR filter, followed by a summation across channels. The filtering operation that produces one output sample of $o(t)$ is obtained by LN multiplications and LN additions.

In this spatio-temporal extension, the covariance matrices in equation 2.3 are replaced with their spatio-temporal generalizations, i.e., $\mathbf{R_s} = E_s\{\tilde{\mathbf{s}}(t)\tilde{\mathbf{s}}(t)^T\}$ and $\mathbf{R_{i+n}} = E_i\{(\tilde{\mathbf{i}}(t) + \tilde{\mathbf{n}}(t))(\tilde{\mathbf{i}}(t) + \tilde{\mathbf{n}}(t))^T\}$. In the remainder of the chapter, we shall always assume the spatio-temporal extension unless otherwise specified, and omit the $\tilde{}$ for notational convenience.

Covariance matrix estimation

In order to compute the optimal filter $\mathbf{w}$, we need to estimate the two covariance matrices $\mathbf{R_s}$ and $\mathbf{R_{i+n}}$ on a training data set, where the training data can be either patient-specific or patient-independent (see subsection 2.2.3).

In our implementation of the algorithm, the seizure covariance matrix $\mathbf{R_s}$ was computed as [2]

$$\mathbf{R_s} \approx \frac{1}{|\mathcal{S}|} \sum_{t \in \mathcal{S}} \mathbf{y}(t)\mathbf{y}(t)^T \tag{2.4}$$

where $\mathcal{S}$ is the set of all samples that are part of epochs that were marked as "seizure" by an expert epileptologist. Similarly, the peak-interference covariance matrix $\mathbf{R_{i+n}}$ was computed as

$$\mathbf{R_{i+n}} \approx \frac{1}{|\mathcal{I}|} \sum_{t \in \mathcal{I}} \mathbf{y}(t)\mathbf{y}(t)^T$$

where $\mathcal{I}$ is the set of all samples that are part of epochs that contain a peak interference. As opposed to $\mathcal{S}$, the set of samples that belong to $\mathcal{I}$ are identified in an automatic fashion. To this end, again, a data-driven filtering technique was used.

The signal was modeled as a linear combination of a seizure time series $\mathbf{s}(t)$ and a seizure-free time series $\mathbf{f}(t)$ (including noise-only and peak-interference-plus-noise segments). A spatial filter $\mathbf{p}$ was then computed as the eigenvector belonging to the largest eigenvalue of $\mathbf{R_f^{-1}R_s}$, where the seizure covariance matrix $\mathbf{R_s}$ was computed as in equation 2.4. The seizure-free covariance matrix $\mathbf{R_f}$ was computed on the rest of the data. Note that in this prior step, we do not use any temporal filtering, i.e. L is set to 1 resulting in a purely spatial filter. This filter was then used to produce a single-channel output given by $\mathbf{p}^T\mathbf{y}(t)$. The RMS over three seconds of the output was computed. The 40 minutes of highest RMS during seizure-free epochs were labeled as peak interference which then form the set $\mathcal{I}$. The set is formed by ranking the signal by highest power, selecting the sample with the highest power, extending the selection to three seconds around the sample and repeating until a total of 40 minutes are selected. This captures most of the peak interferers that appear during a 24-hour recording such as, e.g. chewing artifacts, eye blinking, speech artifacts, head motion, etc while being short enough such that high power interferers dominate

[2] Note that, since the noise floor is always present, the resulting covariance matrix is actually an estimate for $E_s(\mathbf{s}(t)+\mathbf{n}(t))(\mathbf{s}(t)+\mathbf{n}(t))^T = \mathbf{R_s} + \mathbf{R_n}$. If necessary, $\mathbf{R_n}$ can be estimated over noise-only epochs (without seizures or peak interferers) and subtracted from the estimate of $\mathbf{R_s}$. However, since absence seizures have a much higher amplitude than the noise floor generated by background EEG, we have not applied this correction in our implementation.

the set $\mathcal{I}$. The process leading to the identification of the peak interferers and the training of the max-SPIR filter is shown in figure 2.2.

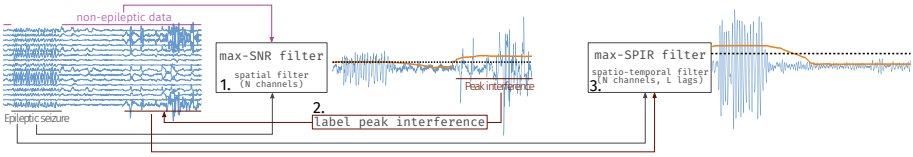


Figure 2.2: Illustration of the three-stage process that is used to train the spatio-temporal max-SPIR filter. (1) Bandpass filtered multichannel EEG (left, blue) is used to train a purely spatial max-SNR filter. The filter optimally amplifies the epochs of epileptic activity and attenuates those of non-epileptic data. (2) Examples of peak interference are automatically labeled based on the filtered data with the highest RMS power during non-epileptic epochs. (3) These peak interferences and the epileptic activity epochs are used to train a max-SPIR spatio-temporal filter. In contrast to the max-SNR filter, the max-SPIR filter will focus on suppressing peak interferers (which cause the majority of the threshold crossings) more so than suppressing sub-threshold noise.

Regularization

The inclusion of time lags substantially increases the dimension of the covariance matrices, possibly making them ill-conditioned due to the redundancy in the entries of $\mathbf{y}(t)$. For this reason a regularization scheme is required. Wouters et al. [83] proposed an effective regularization scheme for a template-matching-filter that optimizes a similar cost function as in equation 2.3. The regularization is obtained by projecting the data on a subspace containing the main principal components of the denominator covariance matrix along with the template itself (to represent the target signal). In this work we adapted the method to our problem formulation. The data were projected on a subspace defined as the span of the principal components of the peak interference $\mathbf{R_{i+n}}$ and seizure $\mathbf{R_s}$ covariance matrices. The principal components with the largest eigenvalues of $\mathbf{R_{i+n}}$ and accounting for 90% of the variance in the interference segments were retained. The principal components with the largest eigenvalues of $\mathbf{R_s}$ accounting for 95% of the variance in the seizure segments were also retained. These two sets of principal component vectors were then combined and orthogonalized by placing them in the columns of a new matrix $\mathbf{M}$ on which a singular value decomposition is applied to find an orthogonal basis for the subspace. Let $\mathbf{U} = [\mathbf{u}_1 \dots \mathbf{u}_K]$ denote the matrix, the columns of which consist

of the K left singular vectors corresponding to the largest singular values of $\mathbf{M}$, where the cut-off K is chosen such that the cumulated sum of these singular values is at least 99% of the sum of all singular values. The matrix $\mathbf{U}$ is then used as a compression matrix on the data, i.e.,

$$\mathbf{y_c}(t) = \mathbf{U}^T\mathbf{y}(t)$$

where $\mathbf{y_c}(t) \in \mathbb{R}^K$. The optimal compressed filter $\mathbf{w_c}$ is then given as the eigenvector of $(\mathbf{R_{i+n,c}^{-1}})\mathbf{R_{s,c}}$, where $\mathbf{R_{i+n,c}} = \mathbf{U}^T\mathbf{R_{i+n}}\mathbf{U}$ and $\mathbf{R_{s,c}} = \mathbf{U}^T\mathbf{R_s}\mathbf{U}$ correspond to the compressed matrices. The filter output is then defined as $o(t) = \mathbf{w_c}^T\mathbf{y_c}(t)$ which is equivalent to an uncompressed filtering $o(t) = \mathbf{w}^T\mathbf{y}(t)$ with $\mathbf{w} = \mathbf{U}\mathbf{w_c}$. It is noted that this compression is only applied with the purpose of regularization during training to obtain a better (uncompressed) filter $\mathbf{w}$. During operation of the algorithm (at test time), we always apply the full filter $\mathbf{w}$ on the uncompressed data, as the compression with $\mathbf{U}$ requires more computations than the filtering with $\mathbf{w}$.

Seizure detection

The power of the single-channel output $o(t)$ was used to detect samples corresponding to a seizure. The RMS of $o(t)$ over three seconds was calculated as this duration is commonly used to define an absence seizure [61]. A threshold was then applied to the running RMS signal, and selected depending on the desired sensitivity (see Section 2.3). This binary classification (above or below threshold) classified samples as seizure or non-seizure.

Several rules were applied to the output of the binary classification to assess the performance of the classification. Samples labeled as seizures and occurring less than 1.5 seconds before the start or after the end of a seizure were not counted as false positive. This was done to account for the settling time of the rolling RMS. False positives occurring less than 30 seconds apart were merged as a single false positive. A seizure was considered as detected if at least one output sample of the binary classification was labeled as a seizure. These rules were applied both to our proposed algorithm and to the algorithms used as benchmark.

2.2.3 Cross-validation

When training and testing the seizure detection algorithm in a patient-specific paradigm, care was taken to separate the dataset in independent training and testing sets. The seizure epochs were split in two folds of equal size (referred to as seizure folds). The rest of the data were split in four folds (referred to as non-seizure folds) of continuous data (each non-seizure fold contained approximately six hours of data). Each training fold used one fold of seizure epochs and three folds of non-seizure data. The obtained filter was then evaluated on the remaining seizure fold and remaining non-seizure fold. This process was repeated for all possible combinations between seizure and non-seizure training folds. An illustration of the cross-validation scheme for the patient-specific case is given in figure 2.3.

Figure 2.3: Patient-specific cross-validation. The seizures are split in two folds. The remaining data are split in four folds. One seizure fold and three non-seizure folds are used for training. The remaining seizure fold and remaining non-seizure fold are used for testing. This process is repeated until all combinations of training and testing folds are covered.

In the patient-independent paradigm, a leave-one-patient-out approach was used for cross-validation. The seizure $\mathbf{R_s}$ and peak interference $\mathbf{R_{i+n}}$ covariance matrices were calculated for each patient, normalized with respect to their trace, and then averaged across patients to obtain the pair ($\mathbf{R_s}$, $\mathbf{R_{i+n}}$) that is used for the filter design in equation 2.3. The process was repeated by systematically leaving one patient out of the training of the filter and testing on that patient.

As a final validation, the patient-independent filter trained on the patients discharged for out-of-clinic monitoring was evaluated on a second dataset with additional patients and a different EEG device, recorded in hospital. This was used as a validation test to show that the algorithm is generalizable to independent EEG recordings.

2.2.4 Channel selection

When evaluating the performance of the algorithm, we investigated the effect of the number of channels (N) on the performance. To this effect a channel selection procedure that selects N out of all available channels was used. It is based on a greedy forward selection. In the first iteration, 18 single-channel SPIR filters are trained and evaluated. The channel with the best performance (in terms of false positives for a sensitivity of 95%) was selected ($N = 1$). In the next iteration, all subsets of two channels that include the previously selected channel were trained and evaluated ($N = 2$). The process was repeated until N channels were selected. Note that the evaluation of the best channels was done on the training fold rather than the test fold. Although this may introduce an overfitting bias to the training fold, the reported performance (on the test fold) is fair in the sense that the channel selection is not optimized with respect to the test fold. The use of a separate validation fold for channel selection was not possible due to the small amount of seizures in some patients.

2.2.5 Evaluation metrics

A varying seizure detection threshold was used to map sensitivity, i.e. the percentage of seizures that are detected by the algorithm, in function of false positive count. This threshold was applied to the running mean of the RMS of the single-channel output $o(t)$ as explained in subsection 2.2.2. A measure of false positive rate was used in place of specificity because seizures are very rare events and specificity of the algorithm does not correctly assess the clinical utility of seizure detection algorithms [84].

In the patient-specific paradigm, the false positives were averaged over the eight different test folds, to calculate an average false positive rate per day.

Because energy efficiency of the algorithm is important when operating in real time on board of a wearable device, the algorithm was also evaluated in terms of memory usage and computing complexity. The computing complexity was calculated as the number of operations (summations and multiplications required by the algorithm). Single operations that do not depend on the number of channels or the number of samples were not counted. Memory was computed with variables stored with 32 bits. These variables could be either integers or floating point.

2.2.6 Comparison with the state of the art

We compared our method with two state-of-the-art algorithms for absence seizure detection using wearable EEG [71], [72]. Kjaer et al.[71] proposed a support-vector machine (SVM) based classifier for use with a single-channel EEG setup. The algorithm extracts many features based both on amplitude and frequency before applying a non-linear SVM classifier. The algorithm suggested by Xanthopoulos et al.[72] is based on a multichannel wavelet-based algorithm followed by a thresholding operation. These algorithms were re-implemented as truthfully as possible based on the original publications and evaluated on our dataset. A more detailed description of the similarities and differences between these algorithms and our proposed method is given in the discussion.

2.3 Results

All results presented in this section are for the main (ambulatory) dataset, unless mentioned otherwise. Data from all but two subjects could be used for analysis. Data from patient 4 were lost due to a recording error. Patient 1 had no seizures during the recording time. These two subjects with no recorded seizures were excluded from the analysis.

2.3.1 Performance analysis

Figure 2.4 shows the false positive rate in function of seizure detection sensitivity when all ($N = 18$) bipolar channels were used and the number of time lags was set to $L = 25$ in a patient-specific paradigm. A distribution of the false positive rate across the subjects is shown for each sensitivity level. The median number of false positives per day for a sensitivity of 95% is 0.5. Note that we report the mean number of false positives across the two seizure folds. This is why the number of false positives per day are reported as multiples of 0.5, despite the fact that we never record longer than a day. In this context, a value of 0.5 false positives per day within a patient means that the filter trained on seizure fold 1 generates 0 false positives per day, while the filter trained on seizure fold 2 generates 1 false positive per day. When displaying results in a boxplot the following conventions are used. The median is represented as a bright square. The box extends from the first to the third quartile. The whiskers extend to 1.5 times the interquartile range ($1.5 * Q3 - Q1$). All points outside this range are represented as large dots.

Table 2.3 shows the number of false positives when detecting all seizures and when detecting all but one seizure in the different patients when all ($N = 18$) bipolar channels were used and the number of time lags was set to $L = 25$ in a patient-specific paradigm. The patient with the greatest number of false positives per day when detecting all seizures has 2.5 false detections per day. The median number of false detections per day is 0.5. The median F_1 score across all patients is 0.95 (best between 100% sensitivity and detection of all but one seizure).

Figure 2.5 shows the number of false positives per day for a sensitivity of 95% when varying both the number of time lags (L) and the number of channels (N) in a patient-specific paradigm. The figure shows a decrease of false positives when the number of channels and/or the number of time lags increase.

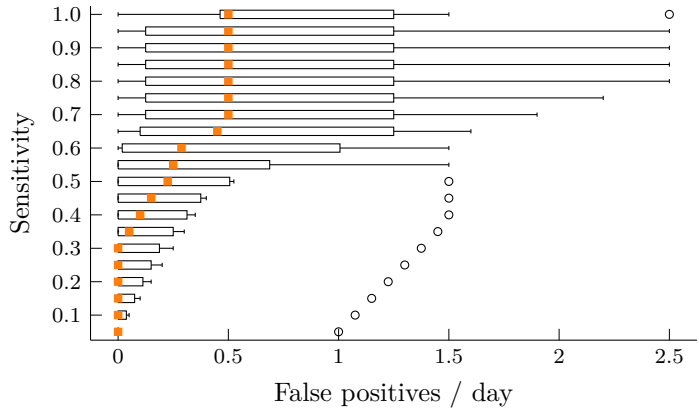


Figure 2.4: Boxplot (over the different subjects) of seizure detection sensitivity as a function of false positives per day when applying the data-driven filter with all ($N = 18$) channels and $L = 25$ time lags in a patient-specific paradigm. Median is represented by an orange square, outliers by a black circle.

Patient	Recorded Seizures	False positives per day for 100% detection	False positives per day to detect all but one seizure	F_1 score
2	41	0	0	1
3	8	0.5	0.5	0.90
5	5	1.5	1.5	0.87
6	5	0.5	0	0.94
7	24	2.5	1	0.96
8	18	0	0	1

Table 2.3: Number of false positives per day for each patient to detect all seizures and to detect all but one seizure when all ($N = 18$) bipolar channels were used and the number of time lags was set to $L = 25$ in a patient-specific paradigm.

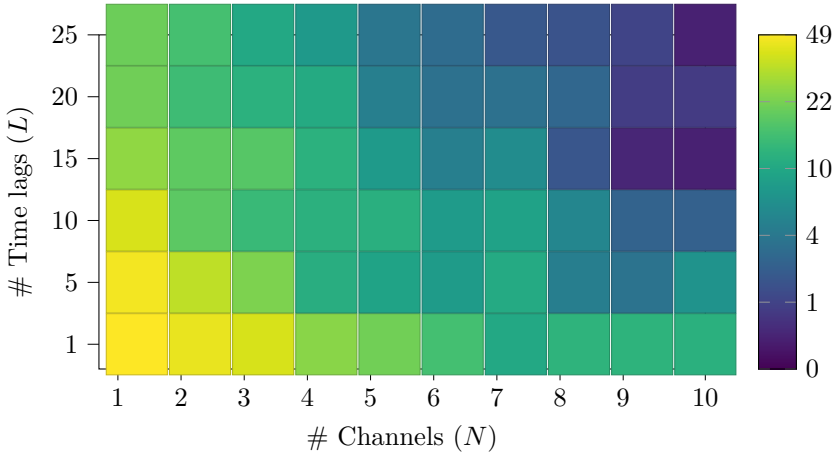


Figure 2.5: Median (over the different subjects) of seizure false positives per day when setting the sensitivity to 95% and varying the number of channels (N) and time lags (L) in a patient-specific paradigm. Color represents false positives per day. Dark blue represents a low number of false detections, yellow a high number.

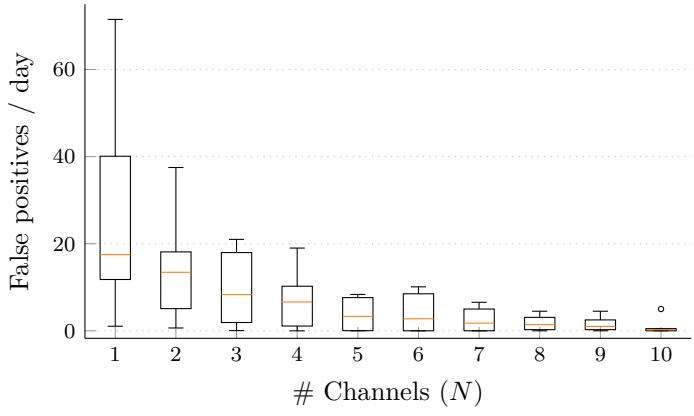


Figure 2.6: Boxplot (over the different subjects) of seizure false positives per day when setting the sensitivity to 95% and varying the number of channels (N) when using $L = 25$ time lags in a patient-specific paradigm.

Figure 2.6 shows how the performance of the algorithm varies in function of the number of channels (N) that are used for the classification task with $L = 25$ time lags in a patient-specific paradigm. The number of false positives per day decreases with increasing number of channels, reaching 0.5 false positives per day when using 10 channels.

Figure 2.7 illustrates the effect of the number of available training seizures on the performance of the algorithm. The algorithm was trained in a patient-specific fashion using all ($N = 18$) channels and $L = 25$ time lags. This graph shows results for three patients (patients 2, 7 and 8), the only ones with 10 or more seizures. Figure 2.7a illustrates that the number of available training seizures has very little influence on the performance of the algorithm. The spatio-temporal pattern in the seizure covariance matrix estimated from one seizure (Fig. 2.7b) is very similar to the seizure covariance matrix estimated from all seizures (Fig. 2.7c).

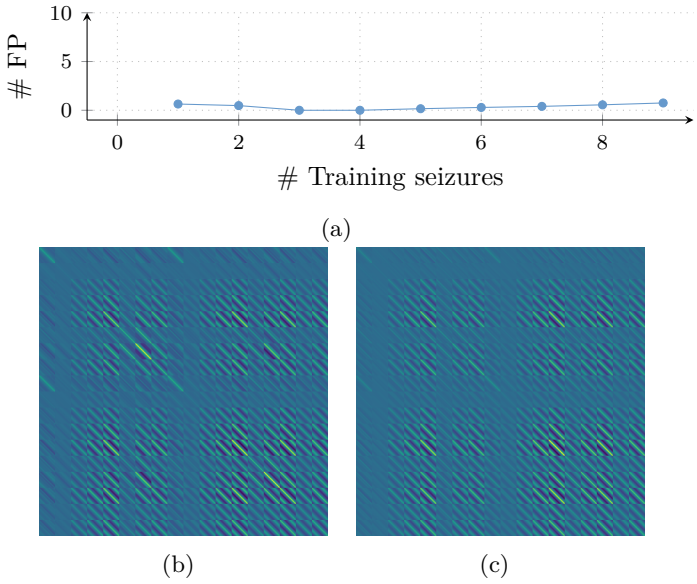


Figure 2.7: (a) Median number of false positives per day for a sensitivity of 95% when artificially limiting the number of training seizures, using all ($N = 18$) channels and $L = 25$ time lags in a patient-specific paradigm. – (b) Example of a seizure covariance matrix estimated from one seizure – (c) Example of a seizure covariance matrix estimated from all seizures in the same patient.

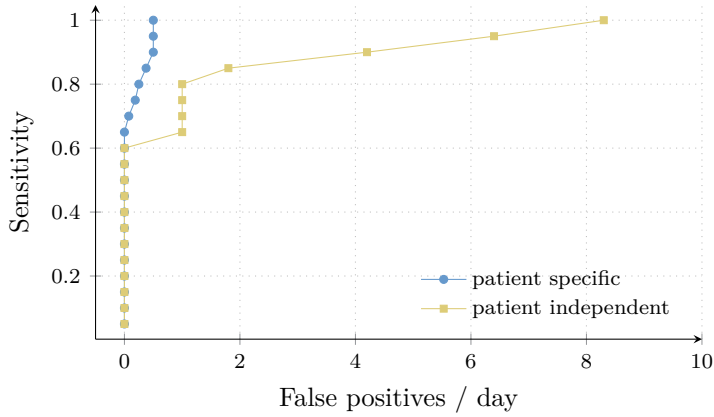
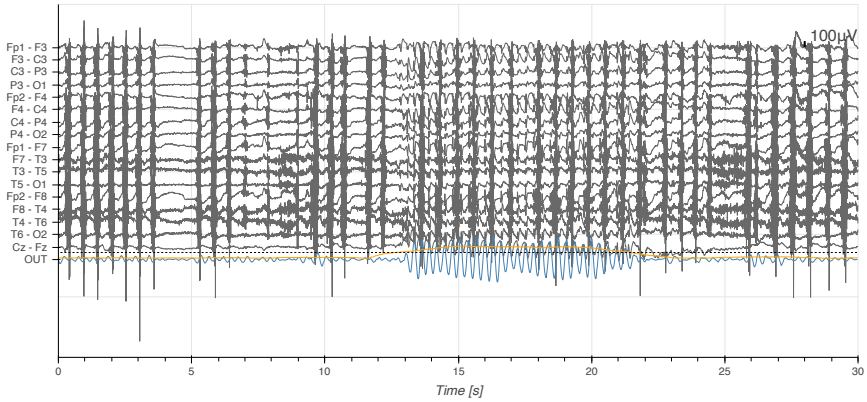


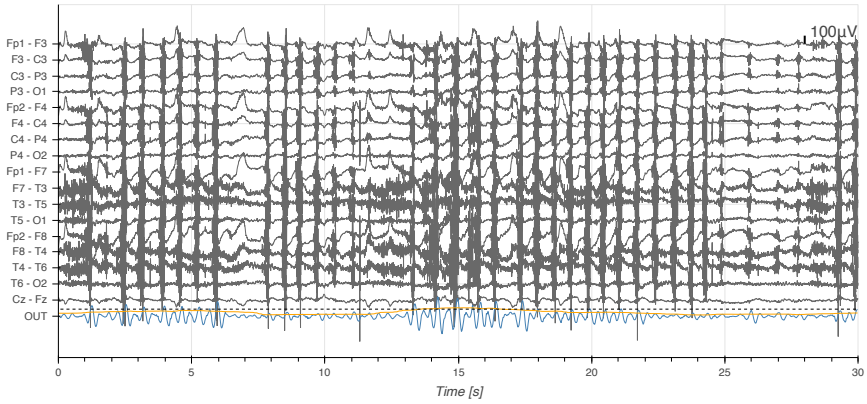
Figure 2.8: Median (over the different subjects) of seizure detection sensitivity in function of false positives per day when applying the data-driven filter with all ($N = 18$) channels and $L = 25$ time lags in a patient-specific paradigm (blue) and a patient-independent paradigm (ocher).

Figure 2.8 compares the performance of the algorithm when trained in a patient-specific paradigm compared to a patient-independent paradigm. Both were generated using all ($N = 18$) channels and $L = 25$ time lags. The sensitivity versus false positive rate starts diverging for sensitivities higher than 60%. For these higher sensitivities the false positive rate per day is higher in the patient-independent paradigm than in the patient-specific one. For a sensitivity of 95% the patient independent-algorithm makes a median of 6 false positives per day while the patient-specific one makes 0.5 false positives per day.

As a final validation, the patient-independent filter trained on the out-of-clinic recordings was evaluated on the validation dataset of patients recorded in hospital in order to test how well the pre-trained filter generalizes to other data sets. For a detection threshold that detects all 222 seizures in the dataset, there were two false detections in two different patients across the whole dataset. An example of a true detection and of a false detection are given in figure 2.9. Both detections occur while the patient is eating. The artifacts resulting from chewing are well suppressed in both examples. The seizure is correctly amplified. In the case of the false detection some noise with a seizure-like pattern triggers a false detection.



(a)



(b)

Figure 2.9: An example of detected seizure (a) and of a false detection (b) in the validation dataset. The blue line is the single-channel output of the filter. The orange line is the running mean of the RMS of the single-channel output. The black dashed line is the detection threshold.

2.3.2 Benchmarking performance and energy consumption

Figure 2.10 shows a comparison of the performance of our proposed algorithm with state-of-the-art algorithms for patient-specific absence seizure detection. Four seizure detection algorithms are shown on the figure. The best performing classifier is a nonlinear SVM classifier on a single-lead EEG [71]. The best single-lead channel was investigated in the work of Kjaer et al. For their algorithm, they found F7 − Fp1 to be the best channel. It has a median false positives per day across subjects of zero for sensitivities up to 85%. For higher sensitivities, the number of false positives increases rapidly. The classification method we propose ($N = 10, L = 25$) performs similarly with 0.25 false positives per day for a sensitivity of 85%. When the method we propose is restricted to one channel ($N = 1, L = 25$), it generates 15 false positives per day. The multi-channel ($N = 18$) wavelet-based algorithm does not perform well on this dataset [72]. It already has 15.5 false positives per day for a sensitivity of 10%. The number of false positives increases with higher sensitivities.

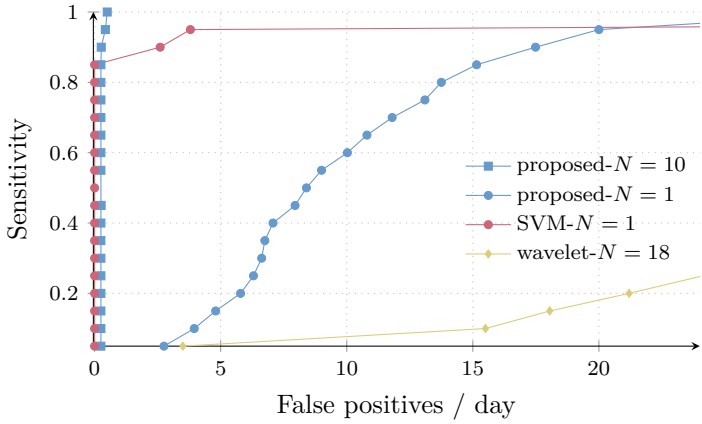


Figure 2.10: Comparison of the sensitivity in function of false positives per day in four different seizure detection algorithms in a patient-specific paradigm. SVM-$N = 1$ is the algorithm of Kjaer et al. [71]. Wavelet-$N = 18$ is the algorithm of Xanthopoulos et al. [72].

The memory usage of the method we propose corresponds to the data that are buffered and in the storage of the filter coefficients. The number of calculations needed to classify are the result of applying the data-driven filter and computing the RMS of the single-channel output. The memory usage of the non-linear SVM classifier on a single-lead EEG [71] resides in a data buffer, the finite

impulse response filter that is used to calculate some features and in the support vectors that need to be stored for classification of a new epoch. The SVM algorithm uses an average of 403 support vectors (full range: [17, 1659]). The number of calculations needed to classify a new epoch are the result of the feature extraction and classification using a radial basis function kernel. The memory usage of the wavelet-based method [72] lies in the data that are buffered. The number of calculations needed to classify a new epoch are the result of the continuous wavelet transform and the transformation to a single-channel output (variance of individual channels and mean over channels). The total number of bytes of memory as well as the number of operations (summations, multiplications, non-linear operations) are reported in table 2.4. The number of operations is reported as a normalized number of operations per sample. For the window-based algorithms, the number of operations on a window is divided by the number of samples in the window. For comparison purposes the sampling frequency of all algorithms was set to 200 Hz. The algorithm we propose uses 100 bytes of memory in the single channel and 1 kB of memory in the 10 channel configuration. The amount of memory required and the number of computations are linearly proportional to the number of channels and the number of time lags (see subsection 2.2.2). The SVM and the wavelet methods use respectively 12 kB and 8 kB of memory. The algorithm we propose executes 50 operations per sample period (25 summations and 25 multiplications) for one channel and 500 operations per sample period for 10 channels. The SVM uses a windowed approach and classifies windows of two seconds. The algorithm executes 6519 operations per sample period, of which 2 are non-linear (cosine and radial basis function). The wavelet algorithm classifies windows of one second. It executes 24012 operations per sample period. Details on the implementation of these algorithms that allows to calculate the number of operations is given in Appendix 2.A.

	Memory [bytes]	Operations per sample
proposed-$N = 10$	1000	500
proposed-$N = 1$	100	50
SVM-$N = 1$	11976	6519
wavelet-$N = 18$	8000	24012

Table 2.4: Memory consumption (bytes) and number of operations of four seizure detection algorithms. proposed-$N = 10$: our proposed method with $N = 10$ channels and $L = 25$ time lags, proposed-$N = 1$: our proposed method with one ($N = 1$) channel and $L = 25$ time lags, SVM-$N = 1$: a single-channel classifier based on a SVM [71], wavelet-$N = 18$: a multichannel ($N = 18$) wavelet-based classifier [72].

2.4 Discussion

In this chapter we propose an absence seizure detection algorithm specifically designed to run on a wearable device. We collected data in real-life, out-of-clinic environments using a portable EEG amplifier. It has a sensitivity of 95% for 0.5 false positives per day when using 10 EEG channels and 25 time lags. The algorithm requires only 1000 bytes of memory and 500 operations per new sample (250 multiplications and 250 summations). This is well within the capabilities of typical microcontrollers that would be found in a wearable EEG device, such as ARM Cortex M4F [66], [67]. The computational efficiency can be further increased by reducing the number of channels or the number of time lags (see Fig. 2.5 where the corresponding trade-off is visualized), the sampling rate, or by reducing the output sample rate of the spatio-temporal filter, thereby avoiding the need to compute a filter output sample for each input sample.

Low power requirements have several advantages for the patient. First, it allows the use of a smaller size power supply, which is relevant considering that batteries are among the larger and heavier components in a wearable device. This facilitates further miniaturization and concealability of wearable EEG sensor devices. Second, requiring less power allows for a longer recording time, which in turn demands less manipulation effort from the patient, who would not need to replace batteries as often as in currently available devices. Furthermore, algorithms with a sufficiently low computational/memory footprint such that the processing can be on board of the device itself allow to bypass the need for a cloud-based processing, thereby eliminating an energy-intensive wireless transfer of the raw data.

A sensitive seizure detection algorithm is a valuable asset to a neurologist in order to have a precise idea of the number of seizures the patient is having. It is well known that patients with absence epilepsy have difficulties keeping an accurate seizure diary. The algorithm we propose could be used by the neurologist to considerably speed up the analysis of EEG recordings. For this use, the algorithm is set to a sensitivity of 100% and the neurologist is only given a handful of events (including some false positives) to review.

We investigated the effect of the number of channels on the performance of the algorithm. The performance of the algorithm greatly improved when more channels were available. Ten well-chosen channels were sufficient to reach 0.5 false positives per day for a sensitivity of 95% in all but one patient. Using three channels was associated with a median of eight false positives per day for a sensitivity of 95%.

We demonstrated that this method requires only a very small number of example

seizure data to train a linear data-driven filter with good performance. Results from the algorithm barely varied when training with one seizure compared to training with more seizures. This was shown both in the performance of the seizure detection algorithm when trained on a low number of seizures and in the visual comparison of the seizure covariance matrices (figure 2.7). This is in contrast with other machine learning approaches such as neural networks or SVMs [71], [85], [86], which require a large amount of training data before being able to classify correctly. The need for very few annotated samples to train our data-driven filter is a major advantage over complex machine learning methods as this allows neurologists to reduce the time they spend annotating EEG files. It also allows for the use of transfer learning techniques where a patient independent version of the filter is refined based on a small set of patient specific examples. While only a small number of seizures are required to train the algorithm, recording quality during these few examples influences the performance of the algorithm. Patient 5 had only five seizures during the recording period. All seizures were short and the recording was of relatively low quality (i.e. noise during the seizure). Patient 5 is the only patient to generate more than 1 false detection per day for a 95% sensitivity.

We compared the performance of our method in a patient-specific paradigm to the same method in a patient-independent paradigm. A patient-specific approach is associated with a lower number of false detections for the same sensitivity as it is able to better capture the specific characteristics of seizures and non-seizure data for that patient. This includes information on the temporal content of seizures and peak interferences as well as spatial information. A patient-independent paradigm allows the algorithm to be used on a patient for whom no previously annotated data are available. This considerably broadens the pool of patients who could potentially benefit from this algorithm. However, the performance of the algorithm deteriorates in the patient-independent paradigm to six false detections per day. A transfer learning approach could be used where a patient-independent filter is used initially, and then updated to a patient-specific filter when more data from the patient become available. This was not explored in this study.

We verified that our algorithm was generalizable to an independent dataset by applying the patient-independent version of the algorithm trained on patients recorded in a out-of-clinic environment to a dataset collected in the hospital. Results from these 17 in-hospital patients showed detection of 222 seizure events collected over more than 14 days of recording which resulted in only two false detections across all patients. The very low number of false positives illustrates that our recordings conducted in a hospital setting contained less artifacts than recordings collected in environments outside the clinic. False detections are usually triggered when the EEG contains high-amplitude noise and when

the spatio-temporal characteristics of this noise happen to overlap with the spatio-temporal characteristics of seizures. As we showed in this study, this is a rare event. We note that the algorithm can be straightforwardly extended with a mechanism that estimates the event duration based on a thresholding of the filter output. In this case, it would be able to distinguish between long epileptiform events and short (typically interictal) epileptiform events. Furthermore, if short interictal events are not of interest to the neurologist, this mechanism can also be used to reduce the number of false positives due to short interfering events. However, since the distinction between ictal and interictal event duration is somewhat arbitrary and because the necessity of it is use-case specific, we did not take event duration into account in our analysis.

We compared our proposed algorithm with two state-of-the-art algorithms for absence seizure detection [71], [72]. Kjaer et al. [71] designed an algorithm for wearable single-lead EEG where data analysis is performed off-line after the recording. The algorithm extracts many EEG features that differentiate an absence seizure from the rest of the EEG signal. A nonlinear SVM classifier is then used to classify two-second EEG epochs. The algorithm performs very well when it is provided with enough training data (more than five seizures). The SVM algorithm cannot be ported directly onto a microcontroller such as found on wearable EEG systems as the calculation of the EEG features, and classification are computationally heavy processes. The single-channel SVM classifier performed similarly to the multichannel version we propose. We also compared our method to a multichannel wavelet-based algorithm developed to detect absence seizures. The algorithm shares some underpinnings with the method we proposed as it is based on a filtering step, reduction from multichannel data to a single-channel output, and a threshold on the single-channel output. However, that algorithm does not use any spatial information on seizures, imposing an equal weight on all channels. That method does not have a data-driven mechanism to suppress peak interferences so that large amplitude artifacts would also trigger the thresholding mechanism leading to false positives. Another difference is that it uses wavelet-based filtering, which is computationally expensive. Xanthopoulos et al. report 4.8 false positives per day for a sensitivity of 80% on a dataset of six people with absence seizures with a total of 40 hours of recording [72]. The algorithm performed less well on our dataset. Xanthopoulos et al. identified that large amplitude artifacts trigger false detections with their method. In our ambulatory EEG recordings, many such high amplitude artifacts are present. These are suppressed by our proposed method by the spatial filtering.

We also built a graphical user interface (GUI) to improve review of EEG recordings from patients with absence seizures. The GUI is presented in appendix 2.B

The sample we used is small, though it appeared to be informative. We also focused on ca. 24h recordings, so that generalization to multiple-day recordings requires further testing. It must be emphasized that the algorithm focused on a single seizure type, which is appropriate in some but not all epilepsy disorders. Absence seizures can be found in many forms of epilepsy across all ages and they are the hallmark seizure type in two epilepsy syndromes, namely childhood absence epilepsy and juvenile absence epilepsy. However, in many epilepsies, they are not the only seizure type occurring in a patient. In addition, other seizure types may have a less characteristic EEG signature and may involve less channels, making accurate detection more challenging [87].

We believe this work is crucial to advance the development of wearable seizure detection systems for long-term monitoring of people with epilepsy outside of the hospital. Additional work is required to further validate and test this algorithm and to actually integrate it in a wearable EEG device. In such a device, the location and distance between the electrodes will be constrained by wearability considerations [88]. This is explored in chapter 4 where the influence inter-electrode distance on an epileptiform activity detector is evaluated. In addition, validation of this method in a larger patient cohort and during long term monitoring is needed before it can be applied in a clinical setting.

2.5 Conclusion

We have proposed a data-driven linear filtering method for absence seizure detection that is designed to run on a microcontroller for use in a wearable EEG-based seizure detection system. The algorithm aims to suppress peak interferers while enhancing seizure signal. It requires only a few annotated seizures to train the optimal filter. Our algorithm has been benchmarked against two state-of-the-art absence seizure detection algorithms (without computation constraints), and was found to perform almost on par with the best of these, while being much more efficient in terms of hardware memory and computational requirements. The algorithm demonstrates the relevance of a generalised eigenvalue decomposition (GEVD) based filtering pipeline for pattern identification in epilepsy. Findings in this chapter also illustrated that reliable detection of epileptic seizures is possible with less than twenty electrodes as found in standard clinical EEG. The next chapter investigates different algorithms for channel selection in GEVD problems.

Appendix

2.A Calculating the number of operations of each algorithm

When counting the operations of the algorithms in the work of Kjaer et al. [71] and Xanthopoulos et al. [72], the following implementation of common algorithms were chosen:

- The fast Fourier transform (FFT) is implemented using the radix-2 Cooley-Tukey algorithm. The algorithm requires $5N\log_2(N)$ operations, where N is the number of samples in a window and is chosen as a power of two [89].
- The continuous wavelet transform was implemented using FFTs. This implementation takes an FFT of the signal, an FFT of the wavelet, multiplies both signals and takes the inverse-FFT [90]. This process is repeated at every scale.

2.B A graphical user interface for seizure review

Review of EEG for the diagnosis of epilepsy is performed by a highly specialised epileptologist. The task requires the annotator to look for epileptiform events in the recording. The epileptologist typically reviews windows of 10 seconds of data at a time. The review is done chronologically, starting from the beginning of the recording until the end. This is a time-consuming task. Analysing a one hour recording requires the review of 360 ten-second long windows; a 24 hour recording requires the review of 8640 ten-second long windows. The epileptologist can also be aided by tools such as a spectrogram of the data that gives an overview of the activity in the full recording in a single overview image, allowing them to concentrate their review on periods with suspicious activity [91]. Some tools also provide an overview plot of more complex features which should better highlight epileptiform activity [92].

We propose a different way to review EEG recordings for the annotation of absence seizures. Instead of reviewing events chronologically, we propose to review them in order of probability of being a seizure event. For this we use the max-SPIR based filter developed in this chapter. The filter is used with fixed weights in a patient-independent mode trained on the data from the home study. This allows use of the algorithm on new data with no prior annotation. This is packaged in a GUI. First, the operator selects an EEG file in the `edf` file format. The file is automatically processed and produces two panels for review of the epileptiform events. The first panel shown in figure 2.B.1 presents an overview of the power of the events detected by the max-SPIR algorithm. The events are sorted by power. A high power of the output of the algorithm indicates a high confidence of the algorithm in the detection of a seizure. The panel allows the user to click on an event to view the waveform. It also counts the number of events above a user-set detection threshold. The second panel shown in figure 2.B.2 gives a detailed view of a detected event. The panel shows 30 seconds of data in the different EEG derivations along with the output of the max-SPIR filter. The events are sorted by detection probability. The panel allows navigating to the next detected event. A typical use of this GUI for annotation of absence seizures would consist of using the overview panel to get an estimate of the number of events in the recording before reviewing the individual events. Individual events are reviewed in order of likelihood. The review would stop once the annotator sees only non-epileptiform events in individual event panel.

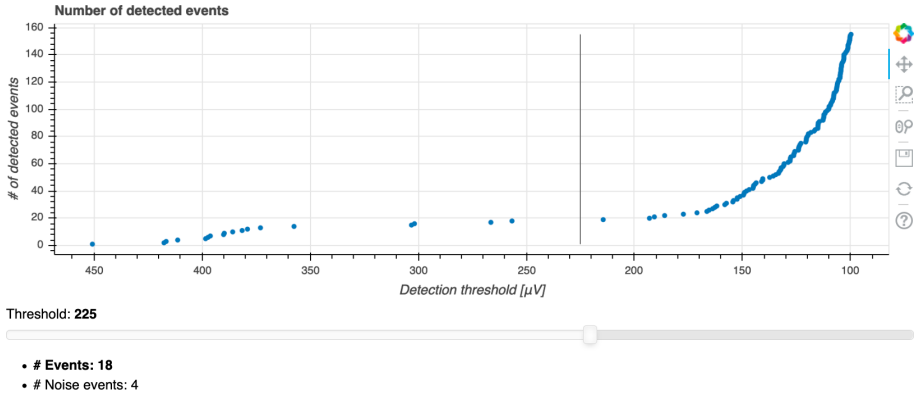


Figure 2.B.1: Overview of the power of the events detected by the max-SPIR algorithm. The events are sorted by power. A high power of the output of the algorithm indicates a high confidence of the algorithm in the detection of a seizure. The panel allows the user to click on an event to view the waveform. It also counts the number of events above a user-set detection threshold.

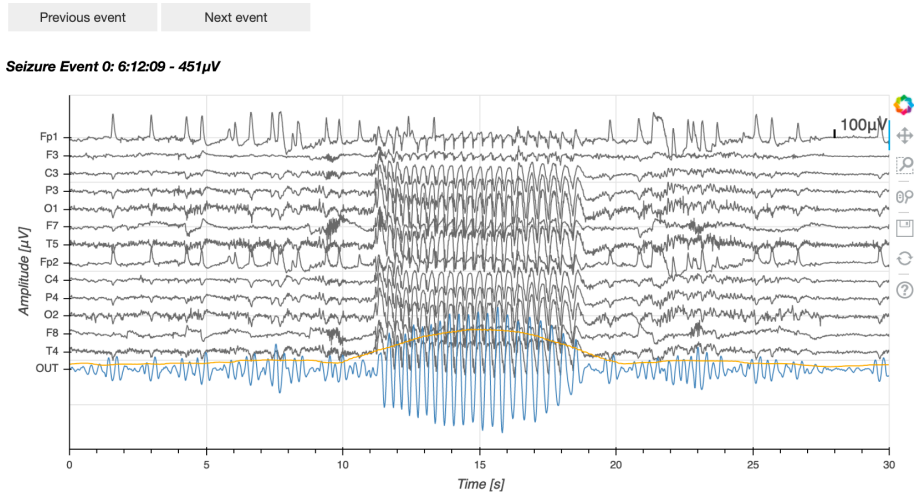


Figure 2.B.2: Detailed view of a detected event. The panel shows 30 seconds of data in the different EEG derivations along with the output of the max-SPIR filter. The events are sorted by detection probability. The panel allows navigating to the next detected event.

CHAPTER 3

Grouped Variable Selection for Generalised Eigenvalue Problems

This chapter is based on J. Dan, S. Geirnaert, and A. Bertrand, "Grouped Variable Selection for Generalised Eigenvalue Problems", *Signal Processing*, vol. 195, 2022. A regularization scheme for the backward elimination method was added in subsection 3.4.5 and the real-world data example (section 3.5) was expanded.

Grouped Variable Selection for Generalised Eigenvalue Problems

Abstract | Many problems require the selection of a subset of variables from a full set of optimization variables. The computational complexity of an exhaustive search over all possible subsets of variables is, however, prohibitively expensive, necessitating more efficient but potentially suboptimal search strategies. We focus on sparse variable selection for generalised Rayleigh quotient optimization and generalised eigenvalue problems. Such problems often arise in the signal processing field, e.g., in the design of optimal data-driven filters. We extend and generalize existing work on convex optimization-based variable selection using semidefinite relaxations toward group-sparse variable selection using the $\ell_{1,\infty}$-norm. This group-sparsity allows, for instance, to perform sensor selection for spatio-temporal (instead of purely spatial) filters, and to select variables based on multiple generalised eigenvectors instead of only the dominant one. Furthermore, we extensively compare our method to state-of-the-art methods for sensor selection for spatio-temporal filter design in a simulated sensor network setting. The results show both the proposed algorithm and backward greedy selection method best approximate the exhaustive solution. However, the backward greedy selection has more specific failure cases, in particular for ill-conditioned covariance matrices. As such, the proposed algorithm is the most robust currently available method for group-sparse variable selection in generalised eigenvalue problems.

3.1 Introduction

Variable selection is an important problem occurring in many mathematical engineering fields. Its goal is to select the subset of variables - often corresponding to specific sensor signals or features thereof - that have the largest impact on the optimization of a specific objective function. Such methods are often used, e.g., to identify the most relevant sensor nodes in a sensor network, or to find the optimal positions to place sensors in a predefined grid [93]. These sensor selection problems arise in many signal processing-related fields, including telecommunication, where antenna placement is critical to the good functioning of a communication network [94]–[101], biomedical sensor arrays, e.g., in the context of EEG channel selection or optimal positioning of wearable sensors [88], [102]–[104], or wireless acoustic sensor networks, where a microphone subset needs to be selected [105], [106]. The number of sensors is typically constrained by practical factors such as fabrication cost, bandwidth, or physical setup limitations, necessitating an appropriate selection of a limited number of sensors and their location.

In many signal processing applications, the objective function can be written as a generalised Rayleigh quotient (GRQ), which corresponds to solving a generalised eigenvalue decomposition (GEVD). Such GRQ- or GEVD-based objectives are encountered in various beamformer or filter design problems, for example, to maximize the signal-to-noise (SNR) [96], [104], [107], [108], or to maximize discriminative properties of the output signals of a filterbank, e.g., in biomedical sensor arrays [109], [110]. In these contexts, variable/sensor selection helps to reduce the computational complexity and power requirements of processing pipelines, to reduce the risk of overfitting of models, and to improve the overall setup.

In this chapter, we focus on grouped variable selection in GRQ/GEVD problems, where the goal is to select a subset of predefined *groups* of variables. For illustrative purposes, but without loss of generality, we will introduce the problem in the context of sensor selection for data-driven spatio-temporal filter design. In sensor networks, an intuitive grouping of the optimization variables is based on the finite impulse response (FIR) filter tap weights in each sensor. However, various (other) types of groupings exist, such as a grouped selection across different filterbands or output filters. Note that all presented methods are besides sensor networks applicable to any other application containing (group-sparse) variable selection for GRQ optimization and GEVD problems.

In sensor selection, the goal is to identify the optimal subset of M out of C available sensors where the choice of M typically leads to a tradeoff between the optimization objective and satisfying some practical constraints. The

exhaustive evaluation of all possible sensor combinations is a computationally costly operation. The selection of M out of C sensors is of combinatorial complexity $\frac{C!}{M!\,(C-M)!}$, where each evaluation requires a new GEVD computation, which in itself has a computational cost of $\mathcal{O}(M^3)$. Therefore, computationally efficient methods are required to solve the sensor selection problem. Two popular heuristic methods are found in the greedy forward selection (FS) and backward elimination (BE) algorithms [102], [111], which are easily applied to many selection problems, including the GEVD problem. However, their greedy nature strongly reduces the combinatorial exploration space, which can result in a highly suboptimal selection. Other approaches take the specific problem structure into account and combine optimization of the objective (e.g., the GRQ) with finding a sparse set of sensors. A specific subclass among these optimization-based approaches relaxes the sensor selection problem to a *convex* optimization problem, which can be solved with off-the-shelf convex optimization solvers [112]. More specifically for the GEVD problem, [108] used the sparsity promoting ℓ_1-norm for a purely spatial beamformer (see Section 3.3.5), which was extended in [94] to a spatio-temporal beamformer using the $\ell_{1,\infty}$-norm as a group-sparse regularizer, albeit in a suboptimal manner (as we will show in Section 3.3.4). Spatial-temporal beamformers allow to combine both spatial and temporal filters in a joint filter design. They have been shown effective in a range of signal processing problems [94], [104], [109]. Other approaches in radar beamforming employed the $\ell_{1,2}$-norm as a group-sparse regularizer in combination with successive convex approximation [99], [100]. Furthermore, [94], [99], [100] only cover the case of a single output filter (multiple-input single-output (MISO) filtering), i.e., a single generalised eigenvector is computed, while several GEVD-based signal processing techniques, such as the common spatial patterns (CSP) filterbank, require the extraction of multiple eigenvectors (multiple-input multiple-output (MIMO) filtering).

The main contributions of this work are as follows:

- We extend the GRQ/GEVD sensor selection for purely spatial filtering in [108] to spatio-*temporal* filtering borrowing techniques from [107]. This necessitates the use of a *group-sparse* regularizer. When a sensor is eliminated, all corresponding filter lags should be put to zero.
- We add the possibility to take multiple filters (i.e., multiple generalised eigenvectors) into account (MIMO), whereas previous work only focused on the dominant generalised eigenvector (MISO). This requires consistent removal or zeroing of the filter coefficients corresponding to an eliminated sensor across *all* filters. This approach can be employed in various other applications, where the notion of a shared selection exists.

- We provide an in-depth and statistical comparison of the proposed method with other state-of-the-art sensor selection methods in GEVD problems, which is largely missing in the aforementioned prior art.

The chapter is structured as follows. First, the sensor selection problem for GEVD and GRQ optimization is introduced in Section 3.2. Next, the convex optimization-based group-sparse sensor selection is explained in Section 3.3. We then thoroughly compare the proposed method with other (benchmark) sensor selection methods on simulated data in Section 3.4. In Section 3.5, we provide an example of the developed method applied on real-world data, in the context of mobile epileptic seizure monitoring. Finally, conclusions are drawn in Section 3.6.

3.1.1 Notation

Scalars, vectors, and matrices are denoted by a lowercase (x), bold lowercase ($\mathbf{x}$), and bold uppercase letter ($\mathbf{X}$). The element of matrix $\mathbf{X}$ on the i^{th} row and j^{th} column is given by x_{ij}. $\mathbf{X}^{\text{T}}$ denotes the transpose of a matrix $\mathbf{X}$ and $\text{Tr}(\mathbf{X})$ denotes the trace of $\mathbf{X}$. The $N \times N$ identity matrix is denoted by $\mathbf{I}_N$, while $\mathbf{0}_N$ denotes an $N \times N$ matrix with zeros. The ℓ_∞-norm of a vector (i.e., the maximal absolute value) is written as $||\mathbf{x}||_\infty$, the ℓ_1-norm of a vector (i.e., the sum of the absolute values) as $||\mathbf{x}||_1$, the ℓ_2-norm of a vector (i.e., the square root of the sum of squared elements) as $||\mathbf{x}||_2$, and the max-norm of a matrix (i.e., the maximal absolute value across all elements) as $||\mathbf{X}||_{\max}$. $\mathbf{X} \succcurlyeq 0$ denotes that $\mathbf{X}$ is a positive semidefinite matrix. The Kronecker-delta is written as δ_{ij} (i.e., $\delta_{ij} = 0$ if $i \neq j$; $\delta_{ij} = 1$ if $i = j$). Finally, the Kronecker-product of matrices $\mathbf{X} \in \mathbb{R}^{I_x \times J_x}$ and $\mathbf{Y} \in \mathbb{R}^{I_y \times J_y}$ is defined as:

$$\mathbf{X} \otimes \mathbf{Y} = \begin{bmatrix} x_{11}\mathbf{Y} & \cdots & x_{1J_x}\mathbf{Y} \\ \vdots & \ddots & \vdots \\ x_{I_x1}\mathbf{Y} & \cdots & x_{I_xJ_x}\mathbf{Y} \end{bmatrix} \in \mathbb{R}^{I_xI_y \times J_xJ_y}.$$

3.2 Sensor selection for GEVD problems

Consider a setting with C sensors and two stationary zero-mean multi-sensor signals $\mathbf{x}_1(t) \in \mathbb{R}^{CL}$ and $\mathbf{x}_2(t) \in \mathbb{R}^{CL}$, where t denotes the sample (time) index and L denotes the group size as explained below. $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ could represent the C sensor signals measured during two different states (e.g., EEG during movement of the left arm and movement of the right

arm [110]), or they could represent two signal components that are both simultaneously present in the sensor signals (e.g., target signal and noise). We assume that the entries of $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ are grouped in blocks of L entries, each group corresponding to a single sensor. For example, a group could consist of L frequency subbands or other features extracted from a single-sensor signal. For illustrative purposes, but without loss of generality, we focus here on spatio-temporal filter design, in which case the L entries of a group correspond to a delay line of length L. In this case, the vector $\mathbf{x}_1(t) \in \mathbb{R}^{CL}$ can be represented as $\mathbf{x}_1(t) = \begin{bmatrix}\underline{\mathbf{x}}_{1,1}(t)^{\mathrm{T}} & \underline{\mathbf{x}}_{1,2}(t)^{\mathrm{T}} & \cdots & \underline{\mathbf{x}}_{1,C}(t)^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$ where $\underline{\mathbf{x}}_{1,c}(t) = \begin{bmatrix}x_{1,c}(t) & x_{1,c}(t-1) & \cdots & x_{1,c}(t-L+1)\end{bmatrix}^{\mathrm{T}}$ represents the causal FIR filter taps corresponding to the c^{th} sensor (similarly for $\mathbf{x}_2(t)$).

The goal is to find a spatio-temporal filter represented by $\mathbf{w} \in \mathbb{R}^{CL}$ which optimally discriminates between the two signals $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$. Optimal discrimination corresponds to maximizing the energy of the output signal $y_1(t) = \mathbf{w}^{\mathrm{T}}\mathbf{x}_1(t)$, while minimizing the energy of the output signal $y_2(t) = \mathbf{w}^{\mathrm{T}}\mathbf{x}_2(t)$. The optimal $\mathbf{w}$ is thus found by maximizing:

$$\max_{\mathbf{w}\in\mathbb{R}^{CL}} \quad \frac{E\{(\mathbf{w}^{\mathrm{T}}\mathbf{x}_1(t))^2\}}{E\{(\mathbf{w}^{\mathrm{T}}\mathbf{x}_2(t))^2\}} = \frac{\mathbf{w}^{\mathrm{T}}\mathbf{R}_1\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\mathbf{R}_2\mathbf{w}}, \tag{3.1}$$

where $\mathbf{R}_1 = E\{\mathbf{x}_1(t)\mathbf{x}_1(t)^{\mathrm{T}}\} \in \mathbb{R}^{CL\times CL}$ and $\mathbf{R}_2 = E\{\mathbf{x}_2(t)\mathbf{x}_2(t)^{\mathrm{T}}\} \in \mathbb{R}^{CL\times CL}$ are the corresponding covariance matrices. Assuming ergodicity and given T samples of the signals $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$, these covariance matrices can be estimated as $\mathbf{R}_1 = E\{\mathbf{x}_1(t)\mathbf{x}_1(t)^{\mathrm{T}}\} \approx \frac{1}{T}\sum_{t=0}^{T-1}\mathbf{x}_1(t)\mathbf{x}_1(t)^{\mathrm{T}}$ and similarly for $\mathbf{R}_2$. The problem in (3.1) is known as a generalised Rayleigh quotient (GRQ) optimization. In the case where $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ represent the target signal and noise components, respectively, Equation (3.1) implies a maximization of the signal-to-noise ratio, resulting in a so-called max-SNR filter [82]. In max-SNR filtering, the covariance matrices $\mathbf{R}_1$ and $\mathbf{R}_2$ thus correspond to the spatio-temporal covariance matrices related to the target signal and the noise, respectively. In the CSP framework [110], these covariance matrices correspond to the two signal classes that have to be discriminated (e.g., left versus right hand movement).

Because of the scale-invariance of $\mathbf{w}$ in (3.1), we can arbitrarily set the output power depicted in the denominator to $\mathbf{w}^{\mathrm{T}}\mathbf{R}_2\mathbf{w} = 1$. Using the method of Lagrange multipliers to solve (3.1) then leads to a GEVD [113]:

$$\mathbf{R}_1\mathbf{w} = \lambda\mathbf{R}_2\mathbf{w}.$$

The optimal filter $\mathbf{w}$ corresponds to the generalised eigenvector (GEVc) corresponding to the largest generalised eigenvalue (GEVl).

In various applications, the GRQ optimization of (3.1) for MISO filtering is generalised to MIMO filtering, i.e., multiple output filters. In this MIMO case, the goal is to find a filterbank of K spatio-temporal filters $\mathbf{W} \in \mathbb{R}^{CL \times K}$ for which the sum of the energies of the multiple output signals is maximally discriminative:

$$\begin{aligned} \max_{\mathbf{W} \in \mathbb{R}^{CL \times K}} \quad & \frac{\mathrm{Tr}\left(\mathbf{W}^{\mathrm{T}} \mathbf{R}_1 \mathbf{W}\right)}{\mathrm{Tr}\left(\mathbf{W}^{\mathrm{T}} \mathbf{R}_2 \mathbf{W}\right)} \\ \text{s.t.} \quad & \mathbf{W}^{\mathrm{T}} \mathbf{R}_2 \mathbf{W} = \mathbf{I}_K, \end{aligned} \tag{3.2}$$

with $\mathrm{Tr}\left(\cdot\right)$ denoting the trace operator and $\mathbf{I}_K$ the $K \times K$ identity matrix. The constraint in (3.2) ensures that the K output channels are orthogonal to each other with respect to the signal component $\mathbf{x}_2(t)$. This constraint is added to obtain K different filters. By plugging this constraint in the cost function in (3.2), we obtain:

$$\begin{aligned} \max_{\mathbf{W} \in \mathbb{R}^{CL \times K}} \quad & \mathrm{Tr}\left(\mathbf{W}^{\mathrm{T}} \mathbf{R}_1 \mathbf{W}\right) \\ \text{s.t.} \quad & \mathbf{W}^{\mathrm{T}} \mathbf{R}_2 \mathbf{W} = \mathbf{I}_K. \end{aligned} \tag{3.3}$$

The solution of (3.3) is now given by taking the K GEVcs corresponding to the K largest GEVls:

$$\mathbf{R}_1 \mathbf{W} = \mathbf{R}_2 \mathbf{W} \mathbf{\Lambda}. \tag{3.4}$$

This generalization to MIMO filters ($K > 1$) is crucial in classification tasks and discriminative analysis as in the CSP framework or in Fisher's discriminant analysis, where the data are projected into a K-dimensional feature space instead of a one-dimensional space. The number of output filters K then introduces a tradeoff between how much information from the original data is preserved and the GRQ (3.2), which becomes smaller (worse) for larger K.

Our goal is to find the optimal subset of $M \leq C$ out of C sensors, with $M \geq K$, for which the ratio of traces in (3.2) is maximal. Note that eliminating a sensor means that all time lags corresponding to that sensor need to be zero. Furthermore, the selected sensors must be consistent across all K filters (i.e., columns of $\mathbf{W}$) to be able to physically select only a few sensors. That is why a *group-sparse* sensor selection is required, i.e., the filter weights are grouped per sensor across time lags and filters, and whole groups are put to zero (i.e., eliminated) rather than the individual elements in a group. In the next section, we present a convex optimization-based approach for this sensor selection problem.

3.3 Optimal group-sparse sensor selection

In this section, we generalize the optimal sensor selection and array design method of [108], which focuses on GRQ optimization and GEVD for purely spatial filtering (i.e., $L = 1$) and for MISO filtering (i.e., $K = 1$), to spatio-temporal filtering and MIMO filtering. Our derivation is based on a similar $\ell_{1,\infty}$-norm regularization technique as proposed in [107] for multicast beamforming and antenna selection. It is noted that during the consolidation of this work, a similar idea to introduce group-sparsity in GEVD problems was published in [94] in the meantime, independently from our work. The work in [94] establishes the $L > 1$ case, yet without generalizing to the $K > 1$ case as also targeted here. Furthermore, our proposed generalization differs from [94] on another crucial aspect, which will be pointed out throughout the derivation (see Section 3.3.4), and which makes that [94] can not be treated as a special case of our proposed general framework. In Section 3.4, we will also empirically compare with [94] for the $K = 1$ setting and demonstrate the superiority of our generalization.

Before pursuing group sparsity in (3.3), let us first vectorize $\mathbf{W} \in \mathbb{R}^{CL\times K}$ as $\mathbf{w} \in \mathbb{R}^{CLK}$, with $\mathbf{w}_k, k \in \{1,\dots,K\}$, the k^{th} spatio-temporal filter:

$$\mathbf{w} = \begin{bmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_K \end{bmatrix}, \mathbf{w}_k = \begin{bmatrix} \mathbf{w}_{k,1} \\ \vdots \\ \mathbf{w}_{k,C} \end{bmatrix}, \text{ and } \mathbf{w}_{k,c} = \begin{bmatrix} w_{k,c,1} \\ \vdots \\ w_{k,c,L} \end{bmatrix}. \tag{3.5}$$

The optimization problem in (3.3) then becomes:

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{R}^{CLK}} \quad & \mathbf{w}^{\mathrm{T}} \left(\mathbf{I}_K \otimes \mathbf{R}_2\right) \mathbf{w} \\ \text{s.t.} \quad & \mathbf{w}_k^{\mathrm{T}} \mathbf{R}_1 \mathbf{w}_{k'} = \delta_{kk'}, \quad \forall k, k' \in \{1,\dots,K\}, \end{aligned} \tag{3.6}$$

with $\otimes$ the Kronecker-product and $\delta_{kk'}$ the Kronecker-delta (i.e., $\delta_{kk'} = 0$, $\forall k \neq k'$; $\delta_{kk'} = 1$, $\forall k = k'$). Notice that we changed the problem in (3.3) to a minimization problem to accommodate for an easy introduction of the sparse regularization term. It can be shown that the solution of (3.3) and (3.6) are the same up to an arbitrary scaling on each $\mathbf{w}_k$, which is irrelevant as generalised eigenvectors are defined up to a scaling. Using the filter-selector matrix $\mathbf{S}_k \in \mathbb{R}^{CL\times CLK}$, where the subscript indicates the selected coefficients of $\mathbf{w}$:

$$\mathbf{S}_k = \begin{matrix} [\end{matrix} \overset{1}{\mathbf{0}_{CL}} \quad \dots \quad \overset{\text{k-1}}{\mathbf{0}_{CL}} \quad \overset{\text{k}}{\mathbf{I}_{CL}} \quad \overset{\text{k+1}}{\mathbf{0}_{CL}} \quad \dots \quad \overset{\text{K}}{\mathbf{0}_{CL}} \begin{matrix}] \end{matrix}$$

with an identity matrix on the k^{th} position to select the k^{th} filter $\mathbf{w}_k$ from $\mathbf{w}$, i.e., $\mathbf{S}_k\mathbf{w} = \mathbf{w}_k$, (3.6) can be rewritten as:

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{R}^{CLK}} \quad & \mathbf{w}^{\mathrm{T}} \left(\mathbf{I}_K \otimes \mathbf{R}_2\right) \mathbf{w} \\ \text{s.t.} \quad & \mathbf{w}^{\mathrm{T}}\mathbf{S}_k^{\mathrm{T}}\mathbf{R}_1\mathbf{S}_{k'}\mathbf{w} = \delta_{kk'}, \quad \forall k, k' \in \{1, \ldots, K\}. \end{aligned} \tag{3.7}$$

3.3.1 Group-sparsity promoting regularization

The goal is now to introduce sparsity in (3.6) on the *sensor* level. This sparsity on the sensor level corresponds to a group-sparse constraint on (3.6), as all lags of all output filters corresponding to a particular sensor need to be set to zero. Therefore, as in [107], we deploy the convex sparsity-promoting $\ell_{1,\infty}$-norm as a proxy for the optimal but non-convex ℓ_0-norm as a regularization term in (3.6).

To simplify the notations in the remainder of the derivations, we define the permutation matrix $\mathbf{P} \in \mathbb{R}^{CLK\times CLK}$ that permutes the elements of $\mathbf{w}$ such that they are first ordered by sensor and then by filter and lags (instead of first by filter as in (3.5)), resulting in $\tilde{\mathbf{w}}$:

$$\tilde{\mathbf{w}} = \mathbf{P}\mathbf{w} = \begin{bmatrix} \tilde{\mathbf{w}}_1 \\ \vdots \\ \tilde{\mathbf{w}}_C \end{bmatrix} \text{ and } \tilde{\mathbf{w}}_c = \begin{bmatrix} \mathbf{w}_{1,c} \\ \vdots \\ \mathbf{w}_{K,c} \end{bmatrix}. \tag{3.8}$$

Using this notation, the $\ell_{1,\infty}$-norm on the sensor level is defined as:

$$||\mathbf{w}||_{1,\infty} = \sum_{c=1}^{C} ||\tilde{\mathbf{w}}_c||_\infty = \sum_{c=1}^{C} \max_{k=1,\ldots,K} ||\mathbf{w}_{k,c}||_\infty , \tag{3.9}$$

where $||\tilde{\mathbf{w}}_c||_\infty$ corresponds to the maximal absolute value across all lags and filters corresponding to sensor c. As the ℓ_1-norm induces sparsity, while the ℓ_∞-norm is only zero when all elements are zero, the $\ell_{1,\infty}$-norm can be used to put groups of coefficients across lags and filters corresponding to one sensor to zero. Furthermore, in [107], it is also shown that any sparsity-inducing norm can be replaced with the squared norm without changing the regularization properties of the problem. Therefore, the sensor selection problem with the

group-sparse regularization term becomes:

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{R}^{CLK}} \quad & \mathbf{w}^{\mathrm{T}}\left(\mathbf{I}_K \otimes \mathbf{R}_2\right)\mathbf{w} + \mu\left(\sum_{c=1}^{C} ||\tilde{\mathbf{w}}_c||_\infty\right)^2 \\ \text{s.t.} \quad & \mathbf{w}^{\mathrm{T}}\mathbf{S}_k^{\mathrm{T}}\mathbf{R}_1\mathbf{S}_{k'}\mathbf{w} = \delta_{kk'}, \quad \forall k, k' \in \{1,\ldots,K\}, \end{aligned} \tag{3.10}$$

where the regularization parameter μ can be used to control the solution's sparsity and thus the number of sensors selected. Note that this is not yet a convex optimization problem due to the quadratic equality constraints.

3.3.2 Semidefinite formulation and relaxation

To transform (3.10) into a convex semidefinite problem (SDP), we are using the following trick, as suggested in [107], [108], [114]:

$$\begin{aligned} \mathbf{w}^{\mathrm{T}}\left(\mathbf{I}_K \otimes \mathbf{R}_2\right)\mathbf{w} &= \mathrm{Tr}\left(\mathbf{w}^{\mathrm{T}}\left(\mathbf{I}_K \otimes \mathbf{R}_2\right)\mathbf{w}\right) \\ &= \mathrm{Tr}\left(\left(\mathbf{I}_K \otimes \mathbf{R}_2\right)\mathbf{w}\mathbf{w}^{\mathrm{T}}\right) \\ &= \mathrm{Tr}\left(\left(\mathbf{I}_K \otimes \mathbf{R}_2\right)\mathbf{V}\right), \end{aligned}$$

where the second equality holds because of the cyclic property of the trace. Per definition, $\mathbf{V} = \mathbf{w}\mathbf{w}^{\mathrm{T}} \in \mathbb{R}^{CLK\times CLK}$ is thus a rank-1 positive semidefinite matrix. Similarly, the equality constraints can be reformulated as:

$$\mathrm{Tr}\left(\mathbf{R}_1\mathbf{S}_{k'}\mathbf{V}\mathbf{S}_k^{\mathrm{T}}\right) = \delta_{kk'}, \forall k, k' \in \{1,\ldots,K\}.$$

Using the following definition of $\tilde{\mathbf{V}}$:

$$\tilde{\mathbf{V}} = \tilde{\mathbf{w}}\tilde{\mathbf{w}}^{\mathrm{T}} = \mathbf{P}\mathbf{V}\mathbf{P}^{\mathrm{T}} = \begin{bmatrix} \tilde{\mathbf{V}}_{11} & \cdots & \tilde{\mathbf{V}}_{1C} \\ \vdots & \ddots & \vdots \\ \tilde{\mathbf{V}}_{C1} & \cdots & \tilde{\mathbf{V}}_{CC} \end{bmatrix},$$

the group-sparse regularization term in (3.10) can be reformulated similarly to [107]:

$$\begin{aligned}\left(\sum_{c=1}^{C}||\tilde{\mathbf{w}}_c||_\infty\right)^2 &= \sum_{c_1=1}^{C}\sum_{c_2=1}^{C}||\tilde{\mathbf{w}}_{c_1}||_\infty\,||\tilde{\mathbf{w}}_{c_2}||_\infty \\ &= \sum_{c_1=1}^{C}\sum_{c_2=1}^{C}\left|\left|\tilde{\mathbf{V}}_{c_1c_2}\right|\right|_{\max} \\ &= \mathrm{Tr}\left(\mathbf{1}_C\mathbf{1}_C^{\mathrm{T}}\mathbf{U}\right), \end{aligned} \tag{3.11}$$

where the max-norm $||\mathbf{A}||_{\max} = \max_{i,j}|a_{ij}|$ is the elementwise maximum over a matrix, $\mathbf{1}_C \in \mathbb{R}^C$ denotes an all-ones vector of length C, and where $\mathbf{U} \in \mathbb{R}^{C\times C}$ is equal to:

$$\mathbf{U} = \begin{bmatrix} \left|\left|\tilde{\mathbf{V}}_{11}\right|\right|_{\max} & \cdots & \left|\left|\tilde{\mathbf{V}}_{1C}\right|\right|_{\max} \\ \vdots & \ddots & \vdots \\ \left|\left|\tilde{\mathbf{V}}_{C1}\right|\right|_{\max} & \cdots & \left|\left|\tilde{\mathbf{V}}_{CC}\right|\right|_{\max} \end{bmatrix}. \tag{3.12}$$

Using the definition of $\mathbf{U}$ in (3.12), we finally obtain:

$$\begin{aligned} \min_{\substack{\mathbf{V}\in\mathbb{R}^{CLK\times CLK},\\ \mathbf{U}\in\mathbb{R}^{C\times C}}} \quad & \mathrm{Tr}\left((\mathbf{I}_K\otimes\mathbf{R}_2)\,\mathbf{V}\right) + \mu\mathrm{Tr}\left(\mathbf{1}_C\mathbf{1}_C^{\mathrm{T}}\mathbf{U}\right) \\ \text{s.t.} \quad & \mathrm{Tr}\left(\mathbf{R}_1\mathbf{S}_{k'}\mathbf{V}\mathbf{S}_k^{\mathrm{T}}\right) = \delta_{kk'}, \forall k,k' \in \{1,\dots,K\}, \\ & \mathbf{U} \geq |\mathbf{S}_{k,l}\mathbf{V}\mathbf{S}_{k',l'}^{\mathrm{T}}|, \forall k,k' \in \{1,\dots,K\}, \\ & \qquad\qquad \text{and } \forall l,l' \in \{1,\dots,L\}, \\ & \mathbf{V} \succcurlyeq 0, \mathrm{rank}(\mathbf{V}) = 1, \end{aligned} \tag{3.13}$$

with the selector-matrix $\mathbf{S}_{k,l} \in \mathbb{R}^{C\times CLK}$ selecting all coefficients across C sensors for a particular filter k and lag l. The second constraint is an elementwise inequality, which ensures that each element of $\mathbf{U}$ (i.e., for each pair of sensors) is larger than the corresponding element for the corresponding pair of sensors across all filters and lags (expressed by the $\forall$ over the filter and lag indices), and thus implements the max-norm operation. The last two constraints ensure the equivalence between $\mathbf{V}$ and $\mathbf{w}\mathbf{w}^{\mathrm{T}}$.

However, (3.13) is still not a convex optimization problem due to the rank-1 constraint. Therefore, we approximate (3.13) by relaxing the rank constraint,

which is a technique known as semidefinite relaxation (SDR) and results in an SDP [114]:

$$\begin{aligned} \min_{\substack{\mathbf{V}\in\mathbb{R}^{CLK\times CLK},\\ \mathbf{U}\in\mathbb{R}^{C\times C}}} \quad & \operatorname{Tr}\left(\left(\mathbf{I}_K \otimes \mathbf{R}_2\right)\mathbf{V}\right) + \mu \operatorname{Tr}\left(\mathbf{1}_C \mathbf{1}_C^{\mathrm{T}} \mathbf{U}\right) \\ \text{s.t.} \quad & \operatorname{Tr}\left(\mathbf{R}_1 \mathbf{S}_{k'} \mathbf{V} \mathbf{S}_k^{\mathrm{T}}\right) = \delta_{kk'}, \forall k, k' \in \{1, \ldots, K\}, \\ & \mathbf{U} \geq |\mathbf{S}_{k,l} \mathbf{V} \mathbf{S}_{k',l'}^{\mathrm{T}}|, \forall k, k' \in \{1, \ldots, K\}, \\ & \qquad\qquad \text{and } \forall l, l' \in \{1, \ldots, L\}, \\ & \mathbf{V} \succcurlyeq 0. \end{aligned} \tag{3.14}$$

This SDR results in practice in a good approximation of the underlying rank-1 solution, potentially using a post-hoc rank-1 approximation of the solution. Note, however, that we are here not interested in the optimal filter coefficients themselves, but only in the selected sensors, which can be retrieved as the non-zero elements of the diagonal of $\mathbf{U}$. Typically, the GEVD problem in (3.4) is afterward recomputed given the selected sensors from (3.14).

3.3.3 Iterative reweighting and algorithm

Similarly to [107], [108], the all-ones matrix $\mathbf{1}_C \mathbf{1}_C^{\mathrm{T}}$ in (3.14) can be replaced with a reweighting matrix $\mathbf{B}^{(i)} \in \mathbb{R}^{C\times C}$ to implement iteratively reweighted ℓ_1-norm regularization [115]. The optimization problem in (3.14) can then be iteratively solved by updating $\mathbf{B}^{(i)}$ as:

$$B_{c_1 c_2}^{(i+1)} = \frac{1}{U_{c_1 c_2}^{(i)} + \epsilon}. \tag{3.15}$$

This iteratively reweighted ℓ_1-norm regularization procedure compensates for the inherent magnitude-dependency of the ℓ_1-norm. Using the ℓ_1-norm as a proxy for the ℓ_0-norm introduces a too large penalty on the elements that have a large magnitude, while it is only relevant to know whether an element is equal to zero or not [115]. The parameter ϵ, which is set as 10% of the standard deviation of the elements of $\mathbf{U}$ without sensor selection (as suggested in [115] and which can be easily computed using the GEVD in (3.4)), avoids division by zero. Initially, $\mathbf{B}^{(1)}$ is set to $\mathbf{1}_C \mathbf{1}_C^{\mathrm{T}}$, i.e., (3.14) is solved. This iterative reweighting procedure generally converges after a few iterations.

To find the optimal set of a specific number M of sensors, a bisection search on the hyperparameter μ of (3.14) can be performed. Once the optimal set of

sensors is found, the corresponding spatio-temporal filters $\mathbf{W}$ can be computed by taking the K GEVcs corresponding to the K largest GEVls of the GEVD in (3.4), using the reduced covariance matrices $\mathbf{R}_{1,2}^{(\text{red})} \in \mathbb{R}^{ML \times ML}$, i.e., by selecting the rows and columns corresponding to the selected sensors. The complete algorithm, which is referred to as GS-$\ell_{1,\infty}$ (GS for group-sparse) in the remainder of the chapter, is summarized in Algorithm 1[1]. The convex optimization problem in (3.14) is solved using the CVX toolbox [116], [117] and MOSEK solver [118].

Remark: It is noted that this algorithm can be easily extended to complex filter coefficients (as often found in beamforming), as the objective function of (3.14) (with the transpose replaced by Hermitian transpose) is a real-valued function, even though it is function of complex variables, while the inequality constraints are also real. This is due to the use of the trace operator in combination with Hermitian (conjugate symmetric) complex-valued matrices.

[1]An open-source toolbox with the MATLAB implementation of this group-sparse sensor selection algorithm can be found online on https://github.com/AlexanderBertrandLab/gsl1infSensorSelection.

Algorithm 1 Group-sparse sensor selection for GEVD (GS-$\ell_{1,\infty}$)

Input:

- $\mathbf{R}_1, \mathbf{R}_2 \in \mathbb{R}^{CL \times CL}$: to-be-discriminated covariance matrices
- M: number of sensors to be selected
- K: number of filters/GEVCs to take into account
- $\mu_{\text{LB}}, \mu_{\text{UB}}$: lower and upper bounds of the bisection search
- $i_{\max}$: maximal number of reweighting iterations

Output: Optimal subset of M sensors and corresponding filters/GEVCs $\mathbf{W} \in \mathbb{R}^{ML \times K}$

1: Define ϵ as 10% of the standard deviation of the elements of $\mathbf{U}$ corresponding to the solution with all sensors (as can be computed from the GEVD in (3.4)) and the tolerance τ as 10% of the minimum across the diagonal of $\mathbf{U}$ corresponding to the solution with all sensors
2: **while** Not M sensors selected **do**
3: Initialize $\mathbf{B}^{(1)} = \mathbf{1}_C \mathbf{1}_C^{\mathrm{T}}$
4: $\mu = \mu_{\text{LB}} + \frac{\mu_{\text{UB}} - \mu_{\text{LB}}}{2}$
5: **while** $\mathbf{U}$ changes and $i \leq i_{\max}$ **do**
6: Solve

$$\begin{aligned} \min_{\mathbf{V} \in \mathbb{R}^{CLK \times CLK}, \mathbf{U} \in \mathbb{R}^{C \times C}} \quad & \operatorname{Tr}\left((\mathbf{I}_K \otimes \mathbf{R}_2)\, \mathbf{V}\right) + \mu \operatorname{Tr}\left(\mathbf{B}^{(i)} \mathbf{U}\right) \\ \text{s.t.} \quad & \operatorname{Tr}\left(\mathbf{R}_1 \mathbf{S}_{k'} \mathbf{V} \mathbf{S}_k^{\mathrm{T}}\right) = \delta_{kk'}, \quad \forall k, k' \in \{1, \ldots, K\}, \\ & \mathbf{U} \geq |\mathbf{S}_{k,l} \mathbf{V} \mathbf{S}_{k',l'}^{\mathrm{T}}|, \quad \forall k, k' \in \{1, \ldots, K\} \text{ and }, \\ & \qquad\qquad\qquad\qquad\quad \forall l, l' \in \{1, \ldots, L\}, \\ & \mathbf{V} \succcurlyeq 0. \end{aligned}$$

7: Update counter i
8: Update $\mathbf{B}^{(i+1)}$ as:

$$B_{c_1 c_2}^{(i+1)} = \frac{1}{U_{c_1 c_2}^{(i)} + \epsilon}$$

9: **end while**

Determine the (number of) sensors $\hat{M}$ selected by comparing the diagonal of $\mathbf{U}$ with the tolerance τ:
  for $c = 1$ to C **do**
    if $U_{cc} > \tau$ **then** c^{th} sensor selected
    else if $U_{cc} < \tau$ **then** c^{th} sensor eliminated
  end for
  Update the regularization parameter bounds as:
  if $\hat{M} > M$ **then** $\mu_{\text{LB}} = \mu$
  else if $\hat{M} < M$ **then** $\mu_{\text{UB}} = \mu$
end while
Compute optimal filters as the K GEVcs corresponding to the K largest GEVls of the GEVD problem with reduced covariance matrices $\mathbf{R}_{1,2}^{(\text{red})} \in \mathbb{R}^{ML \times ML}$:

$$\mathbf{R}_1^{(\text{red})}\mathbf{W} = \mathbf{R}_2^{(\text{red})}\mathbf{W}\mathbf{\Lambda}$$

3.3.4 Special case I: MISO filtering

When taking only one filter into account for the sensor selection (i.e., $K = 1$; MISO filtering), the SDR problem in (3.14) becomes:

$$\begin{aligned} \min_{\substack{\mathbf{V}\in\mathbb{R}^{CL\times CL},\\ \mathbf{U}\in\mathbb{R}^{C\times C}}} \quad & \mathrm{Tr}\left(\mathbf{R}_2\mathbf{V}\right) + \mu\mathrm{Tr}\left(\mathbf{1}_C\mathbf{1}_C^{\mathrm{T}}\mathbf{U}\right) \\ \text{s.t.} \quad & \mathrm{Tr}\left(\mathbf{R}_1\mathbf{V}\right) = 1, \\ & \mathbf{U} \geq |\mathbf{S}_l\mathbf{V}\mathbf{S}_{l'}^{\mathrm{T}}|, \forall l, l' \in \{1,\ldots,L\}, \\ & \mathbf{V} \succcurlyeq 0, \end{aligned} \tag{3.16}$$

with the selector-matrix $\mathbf{S}_l \in \mathbb{R}^{C\times CL}$ selecting all sensor coefficients corresponding to the l^{th} lag. This simplified problem is very similar to the approach proposed in [94], which was independently published during the consolidation of this work. However, the algorithm derived in [94] has a subtle - yet crucial - difference with (3.16) in the inequality constraint $\mathbf{U} \geq |\mathbf{S}_l\mathbf{V}\mathbf{S}_{l'}^{\mathrm{T}}|, \forall l, l' \in \{1,\ldots,L\}$. In [94], a different inequality was proposed, which only takes the diagonal elements of the different blocks of $\mathbf{V}$ into account, i.e., $\mathbf{U} \geq |\mathbf{S}_l\mathbf{V}\mathbf{S}_l^{\mathrm{T}}|, \forall l \in \{1,\ldots,L\}$, while we also take the off-diagonal elements of each block of $\mathbf{V}$ into account (remember: the blocks of $\mathbf{V}$ correspond to sensors when $K = 1$, the elements per block to different combinations of lags (see (3.5))). While leading to fewer inequality constraints and thus resulting in a decreased computational complexity, this relaxation in [94] alters the solution and leads to a suboptimal sensor selection (as empirically shown in Section 3.4). The reason is that the off-diagonal blocks also appear in the first constraint of (3.16), resulting in a mismatch between both constraints. In the remainder of the chapter, the variant of [94] is dubbed 'GS-$\ell_{1,\infty}$-[94]'.

3.3.5 Special case II: purely spatial filtering

In case we do not only constrain to MISO filtering ($K = 1$), but also restrict $\mathbf{w} \in \mathbb{R}^C$ to a purely spatial filter (i.e., $L = 1$), (3.16) is reduced to:

$$\begin{aligned} \min_{\mathbf{V}\in\mathbb{R}^{C\times C},\mathbf{U}\in\mathbb{R}^{C\times C}} \quad & \mathrm{Tr}\left(\mathbf{R}_2\mathbf{V}\right) + \mu\mathrm{Tr}\left(\mathbf{1}_C\mathbf{1}_C^{\mathrm{T}}\mathbf{U}\right) \\ \text{s.t.} \quad & \mathrm{Tr}\left(\mathbf{R}_1\mathbf{V}\right) = 1, \\ & \mathbf{U} \geq |\mathbf{V}|, \\ & \mathbf{V} \succcurlyeq 0, \end{aligned} \tag{3.17}$$

which is equivalent to the approach proposed in [108].

3.3.6 Computational complexity

The computational complexity of the proposed method can be computed from the complexity of interior-point method solvers for quadratic problems with quadratic constraints that are relaxed using semidefinite relaxation. In general, such problems with N^2 variables and T quadratic constraints can be solved to an arbitrary small accuracy ϵ with a complexity of $\mathcal{O}\left(\max(N,T)^4 N^{0.5} \log(\frac{1}{\epsilon})\right)$ [114]. This leads to a complexity of $\mathcal{O}\left((CLK)^{4.5} \log(\frac{1}{\epsilon})\right)$ for our proposed GS-$\ell_{1,\infty}$ algorithm (Algorithm 1).

3.4 Benchmark study

We compare the proposed GS-$\ell_{1,\infty}$ method with other benchmark sensor selection methods on simulated sensor data with known ground-truth[2]. We use the value of the GRQ (3.2) as the performance metric (higher is better).

Besides the exhaustive search, a random search, and the GS-$\ell_{1,\infty}$-[94] method, the proposed GS-$\ell_{1,\infty}$ method is compared with three other sensor selection methods, which are introduced in Section 3.4.1. For the random search, the final GRQ is the mean over 1000 random selections of sensors for a given problem. The setup of the benchmark study is described in Section 3.4.2. The aforementioned methods are compared using only one filter (MISO filtering) in Section 3.4.3 and using multiple filters (MIMO filtering) in Section 3.4.4 (for those methods that allow for $K > 1$). Finally, we provide a more in-depth comparison of the two best-performing algorithms, namely the proposed GS-$\ell_{1,\infty}$ method and the backward greedy elimination method (see Section 3.4.1) in Section 3.4.5.

[2]We provide an open-source MATLAB implementation of the benchmark study online on https://github.com/AlexanderBertrandLab/benchmarkStudySensorSelection.

3.4.1 Benchmark sensor selection methods

In this section, we briefly introduce other algorithms for sensor selection that will be included in the benchmark study.

Greedy sensor selection methods

Greedy sensor selection methods - also dubbed 'wrapper' methods [102] - sequentially select or eliminate those sensors that maximally increase or minimally decrease the objective, respectively. While these greedy approaches are computationally more efficient than the method proposed in Section 3.3, due to their sequential nature, the greedy mechanism can result in suboptimal selections, as they are stuck with the selected or eliminated sensors from previous steps. The computational complexity of these greedy methods is dominated by the GEVD computation performed at each iteration, which is $\mathcal{O}\left((CLK)^3\right)$ [119]. The greedy selection can be applied in two directions (forward or backward):

Forward selection (FS) The FS method starts from an empty set of sensors and sequentially adds the sensor (i.e., group of KL variables) that maximally increases the objective (3.2). New sensors are added until M out of C sensors are selected.

Backward elimination (BE) The BE method starts from the full set of sensors and sequentially removes the sensor that minimally decreases the objective (i.e., the objective in (3.2)) until M out of C sensors are selected. Many variations on the FS and BE method exist, mostly presented in the context of feature selection for classification [111].

The STECS method

We also compare with the spatio-temporal-filtering-based channel selection (STECS) approach proposed for the GEVD problem in [120]. In the STECS method, initially proposed for $K = 1$, the following optimization problem is solved [120] as a regularized proxy for (3.1):

$$\min_{\mathbf{w}\in\mathbb{R}^{CL}} \quad \mathbf{w}^{\mathrm{T}}\mathbf{R}_2\mathbf{w} + \frac{1}{\mathbf{w}^{\mathrm{T}}\mathbf{R}_1\mathbf{w}} + \mu\,||\mathbf{w}||_{1,2}\,, \tag{3.18}$$

with the $\ell_{1,2}$-norm, defined as $||\mathbf{w}||_{1,2} = \sum_{c=1}^{C} ||\mathbf{w}_c||_2$, enforcing the group sparsity over different lags. The GRQ in (3.1) is split into the first two terms of (3.18) as it removes the scale-invariance of $\mathbf{w}$, while yielding an equivalent solution [120].

The c-th sensor is then selected if $||\mathbf{w}_c||_2$ is larger than a predefined tolerance τ, which is set as 10% of the minimum $\min_{c\in\{1,\dots,C\}} ||\mathbf{w}_c||_2$ of the solution with all sensors. The optimization problem in (3.18) is solved using the line-search method proposed in [120]. The same settings as in [120] have been used. Similarly to the proposed method in Section 3.3.3, we use a bisection search method on the regularization parameter μ to obtain the correct number of sensors M. Furthermore, the authors propose to use the selected sensors with (3.18) for the first filter $\mathbf{w} \in \mathbb{R}^{CL}$ to recompute the solution of $\mathbf{W} \in \mathbb{R}^{CL\times K}$ for multiple filters with (3.4), which means the selection does not take the full objective (3.2) into account. Lastly, note that the optimization problem (3.18) is non-convex, resulting in potential convergence to non-optimal local minima.

Other sensor selection methods for GEVD problems (and in particular in biomedical applications for CSP problems) have been proposed as well [102], for example, based on filter coefficients magnitude (e.g., [121]), or other variants of ℓ_1-norm regularization (e.g., [99], [100], [122], [123]). However, we do not further consider these methods, as they are not designed for group-sparsity and/or the MIMO case, or have been shown to be outperformed by at least one of the aforementioned methods [120].

3.4.2 Setup

Simulation model

We assume a $\sqrt{C} \times \sqrt{C}$ square grid of C sensors, each of which are measuring a mixture of N_1 source signals to be maximized (i.e., contributing to $\mathbf{x}_1(t)$ and the numerator of (3.2)) and N_2 source signals to be minimized (i.e., contributing to $\mathbf{x}_2(t)$ and the denominator of (3.2)) as well as independent sensor noise. This simulated problem resembles point-source models as, for example, found in sensor networks, microphone arrays, neural activity (EEG), and telecommunications. The source signals contributing to $\mathbf{x}_2(t)$ have a power that is approximately 150 times larger than the source signals contributing to $\mathbf{x}_1(t)$. An example is given in Figure 3.4.1. In the max-SNR filtering case, one could think of the source signals contributing to $\mathbf{x}_1(t)$ as target signals and source signals contributing to $\mathbf{x}_2(t)$ as noise signals. In that case, the GRQ can be interpreted as an SNR.

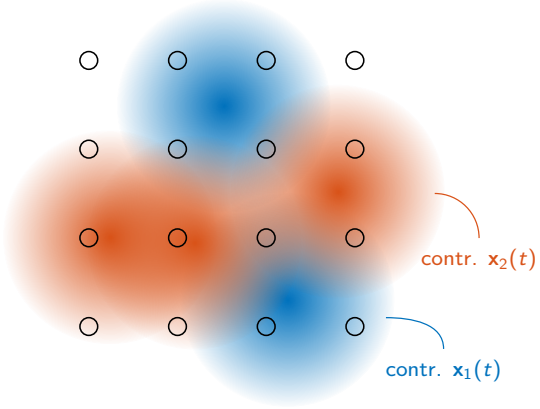


Figure 3.4.1: An exemplary generated problem with $C = 16 = 4 \times 4$ sensors, $N_1 = 2$ random signals contributing to $\mathbf{x}_1(t)$, and $N_2 = 3$ random signals contributing to $\mathbf{x}_2(t)$. Each sensor measures a mixture of the underlying sources. The brightness of the color represents the intensity of the signal as perceived by a sensor.

Each source signal is a bandpass-filtered white Gaussian signal in a random frequency band between 1 and 9 Hz, sampled at 20 Hz. It originates from a random location within the grid of sensors (drawn from a uniform distribution over the entire area) and propagates with an exponentially decaying amplitude to the sensors. The spread of the exponential decay is set such that the maximal attenuation is equal to a predefined attenuation of 0.5%. Furthermore, a source signal is measured at each sensor with a time delay linear to the distance to that source signal, such that the maximal delay is 100 ms (i.e., 2 samples). The sensor noise at each sensor is white Gaussian noise with twice the maximal attenuation as amplitude.

Monte Carlo runs

For each experiment, i.e., for a given number of sensors C, number of lags L, and number of filters K, 250 (for the $K = 1$ case) and 100 (for the $K > 1$ case) of the random problems in Section 3.4.2 are generated, and the results for each evaluated sensor selection method are averaged over these different problems. For each of these Monte Carlo runs, unless specified otherwise, the number of signals N_1 and N_2 is randomized between 1 and $2C$.

	μ_{LB}	μ_{UB}	$i_{\max}$	max. it. bisection search
GS-$\ell_{1,\infty}$	10^{-5}	100	15	20
GS-$\ell_{1,\infty}$-[94]	10^{-5}	10^{4}	15	20
STECS	0	10^{16}	/	200

Table 3.4.1: The chosen hyperparameters of the different optimization-based sensor selection methods.

Hyperparameter choice

Table 3.4.1 shows the chosen hyperparameters for the different optimization-based sensor selection methods. The bisection search for the GS-$\ell_{1,\infty}$ (Algorithm 1), the GS-$\ell_{1,\infty}$-[94], and the STECS method is aborted if no solution was found after a certain number of iterations. For STECS, this number is taken much larger, which is possible due to its computational efficiency. However, this early stopping criterion leads to a limited number of cases where no solution is found for a certain M. To still produce a meaningful solution in those cases, a random extra sensor is added to the solution obtained for $M-1$ sensors, and the corresponding output GRQ (in dB) is computed. However, when no solution is found for the lowest value of M and the previous solution to still produce a meaningful solution correspondingly fails (because it relies on the solution of the lowest M), the results for all methods for those M for which there is no solution in that particular run are removed.

Furthermore, the hyperparameter μ for the GS-$\ell_{1,\infty}$ (Algorithm 1) and GS-$\ell_{1,\infty}$-[94] algorithm is defined relative to the first target objective part of (3.16) (i.e., $\text{Tr}\,(\mathbf{R}_2\mathbf{V})$) for the solution with all sensors.

Statistical comparison

To identify statistically significant differences based on hypothesis testing, we use a linear mixed-effect model (LMEM) [124]. Such an LMEM allows the exploitation of all structure in the data by modeling the obtained GRQ as a function of the method while taking the variation due to the different Monte Carlo runs and the effect of a different number of selected sensors M into account as random factors. The following LMEM is chosen based on the Akaike information criterion (AIC), which takes the model fit and complexity into account:

$$\text{GRQ} \sim 1 + \text{method} + (M|\text{run}).$$

This notation is often used in LMEMs to reflect that the GRQ is modeled with the method as a fixed effect, the number of selected sensors M as random slope (i.e., the GRQ can vary as a function of M independently for each method), and the run as a random intercept. Per fixed term, the estimated regression coefficients (β), standard error (SE), degrees of freedom (DF), t-value, and p-value are reported. If a significant effect between the different methods is found, we use an additional Tukey-adjusted post-hoc test to assess the pairwise differences between individual methods. The significance level is set to $\alpha = 0.05$. All statistical analyses are performed using the R software package and the `nlme` and `emmeans` packages.

In the statistical hypothesis testing, we limit the number of selected sensors to $\frac{C}{2}$, as we consider this lower half range much more relevant in the context of sensor selection than the upper half range. Typically, one wants to drastically reduce the number of required sensors, i.e., below half of the number of available sensors.

3.4.3 Comparison in the MISO case ($K = 1$)

First, we evaluate and compare the presented methods in the first special MISO case of Section 3.3.4, where only one filter (first GEVc) is taken into account, i.e., $K = 1$. In this case, we can also include the comparison with GS-$\ell_{1,\infty}$-[94], which was designed specifically for this case. We choose $C = 25, L = 3$ and look for the optimal sensor selection for M ranging from 2 to 24. Figure 3.4.2 shows the output GRQs as a function of M for each separate method (mean over 250 Monte Carlo runs $\pm$ the standard error on the mean). Table 3.4.2 shows the outcome of the statistical analysis, for M ranging from 2 to 12 (see Section 3.4.2). All presented methods achieve significantly higher (better) GRQ than random selection but lower (worse) GRQ than the optimal solution obtained through an exhaustive search over all possible combinations (Table 3.4.2b).

The greedy sensor selection methods suffer from intrinsic limitations, i.e., they depend on previous choices in their sequential procedure. For example, the FS method starts with a GRQ close to optimal but diverges from the optimal exhaustive solution when M increases, and the other way around for the BE method. However, the FS method achieves overall lower GRQ than the BE method, which is also confirmed by the statistical testing in Table 3.4.2b. This could be due to the fact the FS method is limited to selecting one sensor at a time, which hampers its capacity to probe combined effects of multiple sensors.

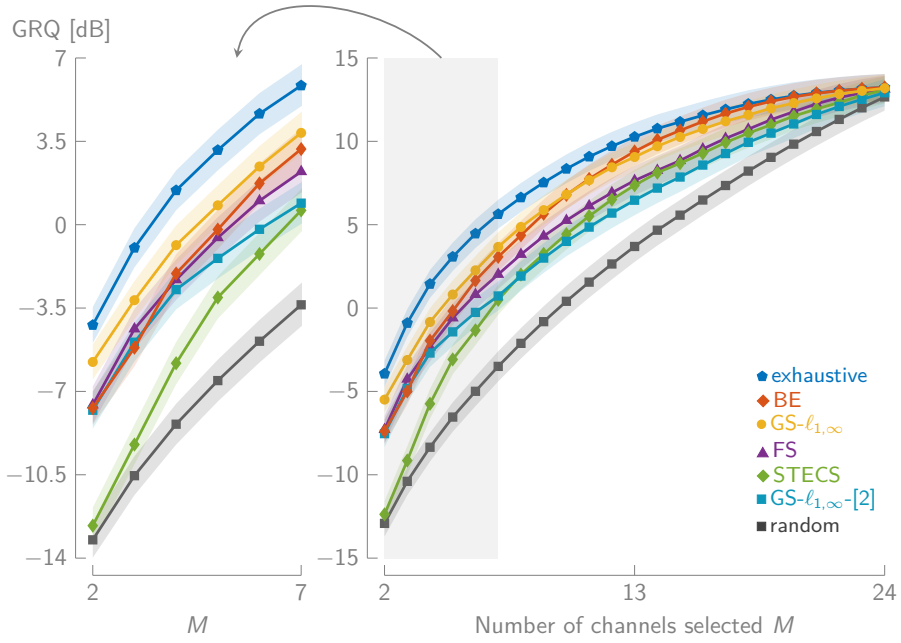


Figure 3.4.2: The output GRQ (mean over 250 runs) as a function of M for the different sensor selection methods when $C = 25, L = 3, K = 1$. The shading represents the standard error on the mean.

Furthermore, Figure 3.4.2 and Table 3.4.2b show that the GS-$\ell_{1,\infty}$-[94] method is outperformed by all other methods, except by the STECS method. Our proposed method (significantly) outperforms GS-$\ell_{1,\infty}$-[94] across all M. This is an effect of dropping the off-diagonal blocks of the inequality constraints of (3.16). However, the gap between both methods becomes smaller for lower M (see also Figure 3.4.2). Similarly, the STECS method is outperformed, especially for low M, by all other methods, suffering from its non-convex objective function. This method achieves slightly higher GRQ than the GS-$\ell_{1,\infty}$-[94] method for most larger M, but it achieves much lower GRQ for small M. As a result, there is also a significant difference observed across all M between 2 and 12 between the STECS and GS-$\ell_{1,\infty}$-[94] method (Table 3.4.2b).

A remarkable conclusion is that the greedy BE method significantly outperforms almost all other state-of-the-art methods, including GS-$\ell_{1,\infty}$-[94] and STECS, which have not been benchmarked in a group-sparse setting against BE in the corresponding original papers [94] and [120], respectively. The only method that

significantly outperforms the BE method is our proposed GS-$\ell_{1,\infty}$ algorithm. Interestingly, although the BE method seems to perform slightly better for larger M, the GS-$\ell_{1,\infty}$ method seems to perform better than the BE method for *small* M especially, explaining the statistically significant difference. The gap between both methods is also larger for these small M than for large M. From an application-based point of view, these smaller M - below half of the total number of sensors C - are often targeted in practice. Indeed, sensor selection is typically performed to substantially decrease the number of required sensors, not to remove only a few sensors. Although the heuristic BE method is computationally much more efficient than the optimization-based GS-$\ell_{1,\infty}$ method, it thus performs worse than the GS-$\ell_{1,\infty}$ method when it most matters. Lastly, it is interesting to identify in how many and in which cases a sensor selection method completely fails. For example, one could define a failure as more than 10 dB difference with the exhaustive method. Using this rule, the fail rate for the BE method across all runs and again for M between 2 and 12 is 3.71%, while this is 0.44% for the GS-$\ell_{1,\infty}$ method. Thus, the BE method has almost 10 times more severe fail cases than the GS-$\ell_{1,\infty}$ method. While these percentages might seem marginal at first sight, it should be taken into account that this percentage is biased by the highly randomized simulated scenarios. After a closer look, these fail cases turn out to mainly occur in cases where the covariance matrix $\mathbf{R}_2$ is ill-conditioned, which is not necessarily a rare case in practical settings, for example, as found in miniaturized EEG sensor networks [88]. In Section 3.4.5, we further analyze these particular fail cases and provide a more extensive discussion.

Fixed-effect term	β	SE	DF	t-value	p-value
intercept	-5.38	0.79	18729	-6.80	< 0.0001
method = exh$-$BE	2.51	0.07		34.94	< 0.0001
method = exh$-$FS	3.35			46.71	< 0.0001
method = exh$-$STECS	5.48			76.30	< 0.0001
method = exh$-$random	8.73			121.59	< 0.0001
method = exh$-$GS-$\ell_{1,\infty}$-[94]	4.33			60.36	< 0.0001
method = exh$-$GS-$\ell_{1,\infty}$	1.83			25.46	< 0.0001

(a)

	exh	BE	FS	STECS	rand	GS-H	GS
exh	/	2.51*	3.53*	5.48*	8.73*	4.33*	1.83*
BE	-2.51*	/	0.85*	2.97*	6.22*	1.83*	-0.68*
FS	-3.53*	-0.85*	/	2.13*	5.38*	0.98*	-1.53*
STECS	-5.48*	-2.97*	-2.13*	/	3.25*	-1.15*	-3.65*
rand	-8.73*	-6.22*	-5.38*	-3.25*	/	-4.40*	-6.90*
GS-H	-4.33*	-1.83*	-0.98*	1.15*	4.40*	/	-2.51*
GS	-1.83*	0.68*	1.53*	3.65*	6.90*	2.51*	/

(b)

Table 3.4.2: (a) The LMEM fixed-effect outcomes for $M = 2$ to 12 when $C = 25, L = 3, K = 1$. (b) The pairwise differences, showing the estimated difference between average GRQ (method in row $-$ method in column)/p-value per pair of methods (p-values < 0.0001 are indicated with $*$). Statistically significant differences are color coded. Values in green/red indicate that the method in the row outperforms/is outperformed by the method in the column. The method names are abbreviated as **exh**: exhaustive, **BE**: Background Elimination, **FS**: Forward Selection, **STECS**, **rand**: random, **GS-H**: GS-$\ell_{1,\infty}$-[94], **GS**: GS-$\ell_{1,\infty}$

3.4.4 Comparison in the MIMO case ($K > 1$)

Figure 3.4.3 and Table 3.4.3a show the results of 100 Monte Carlo simulations with $C = 25, L = 2$, and $K = 2$, i.e., the more general case where $K > 1$. As the GS-$\ell_{1,\infty}$-[94] method was only proposed for $K = 1$, it is not included in these simulations.

First of all, the results confirm that the proposed extension to MIMO filtering in Section 3.3 is valid, as the GS-$\ell_{1,\infty}$ method still obtains GRQs close to those of the optimal exhaustive solution. Furthermore, the results are in line with Section 3.4.3. The BE and GS-$\ell_{1,\infty}$ method again show a statistically significant difference when evaluated across M between 2 and 12 (Table 3.4.3b), confirming that the latter has the advantage for the more relevant low M. Finally, both methods significantly outperform the other benchmark methods (except the exhaustive search).

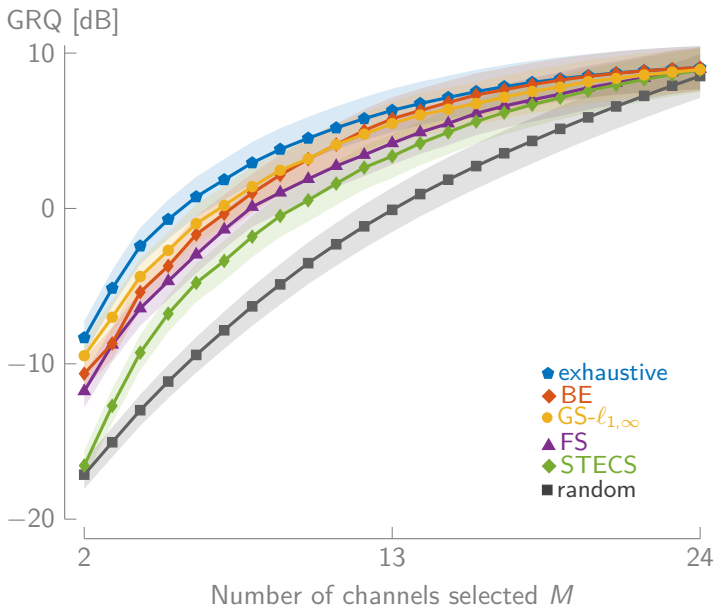


Figure 3.4.3: The output GRQ (mean over 100 runs) as a function of M for the different sensor selection methods when $C = 25, L = 2, K = 2$. The shading represents the standard error on the mean.

Fixed-effect term	β	SE	DF	t-value	p-value
intercept	-14.64	0.91	6393	-16.01	< 0.0001
method = exhaustive$-$BE	2.10	0.11		18.54	< 0.0001
method = exhaustive$-$FS	3.18			28.04	< 0.0001
method = exhaustive$-$STECS	5.35			47.15	< 0.0001
method = exhaustive$-$random	9.09			80.17	< 0.0001
method = exhaustive$-$GS-$\ell_{1,\infty}$	1.51			13.36	< 0.0001

(a)

	exhaustive	BE	FS	STECS	random	GS-$\ell_{1,\infty}$
exhaustive	/	2.10^*	3.18^*	5.35^*	9.09^*	1.51^*
BE	-2.10^*	/	1.08^*	3.24^*	6.99^*	-0.59^*
FS	-3.18^*	-1.08^*	/	2.17^*	5.91^*	-1.66^*
STECS	-5.35^*	-3.24^*	-2.17^*	/	3.75^*	-3.83^*
random	-9.09^*	-6.99^*	-5.91^*	-3.75^*	/	-7.58^*
GS-$\ell_{1,\infty}$	-1.51^*	0.59^*	1.66^*	3.83^*	7.58^*	/

(b)

Table 3.4.3: (a) The LMEM fixed-effect outcomes for $M = 2$ to 12 when $C = 25, L = 2, K = 2$ (GS-$\ell_{1,\infty}$-[94] is omitted as it is only defined for $K = 1$). (b) The pairwise differences, showing the estimated difference between average GRQ (method in row $-$ method in column)/p-value per pair of methods (p-values < 0.0001 are indicated with $*$). Statistically significant differences are color coded. Values in green/red indicate that the method in the row outperforms/is outperformed by the method in the column.

3.4.5 Comparison of GS-$\ell_{1,\infty}$ with BE

In this section, we zoom in on the comparison between the BE and GS-$\ell_{1,\infty}$ method, as these two methods achieve the highest GRQ in the previous simulations. As explained in Section 3.4.1, the BE method can suffer from its greedy sequential selection, where previously eliminated sensors can not be recovered when selecting a lower number of sensors. This inherent disadvantage of the BE method can lead to various fail cases (defined here as $> 10\,\text{dB}$ difference with the optimal exhaustive solution). After closer inspection, we identified that the majority of the fail cases (73.53% of all the fail cases in the previous simulations) corresponded to scenarios in which the matrix $\mathbf{R}_2$ was ill-conditioned, i.e., where there was a large difference between the largest and smallest eigenvalue(s).

To thoroughly test this case, we compare the BE method to the GS-$\ell_{1,\infty}$ method on the subset of 60 simulations of Section 3.4.3 where $2 \leq N_2 \leq 12$, i.e., where the number of signals contributing to $\mathbf{x}_2(t)$ is less than half of the $C = 25$ sensors. These cases correspond to ill-conditioned covariance matrices $\mathbf{R}_2$ in the denominator of the GRQ, where the smallest eigenvalues are determined solely by white Gaussian sensor noise (see Section 3.4.2). The results are shown in Figure 3.4.4, where also the performance of the exhaustive solution is shown as a reference. For large M, both methods perform similarly to the exhaustive solution, with very little change in GRQ for increasing values of M. When M decreases, both methods start to diverge from the exhaustive solution. However, the BE method achieves lower GRQ than the GS-$\ell_{1,\infty}$ method for smaller M. This is confirmed by the LMEM including only those two methods, as there is again a significant effect of the method, i.e., the GS-$\ell_{1,\infty}$ method outperforms the BE method (Table 3.4.4).

Fixed-effect term	β	**SE**	**DF**	**t-value**	**p-value**
intercept	31.30	0.44	1259	71.47	< 0.0001
method = GS-$\ell_{1,\infty}$−BE	1.53	0.19	1259	8.20	< 0.0001

Table 3.4.4: The LMEM outcomes when including only the BE and GS-$\ell_{1,\infty}$ methods in the case with an ill-conditioned covariance matrix $\mathbf{R}_2$ in the denominator (for $M = 2$ to 12).

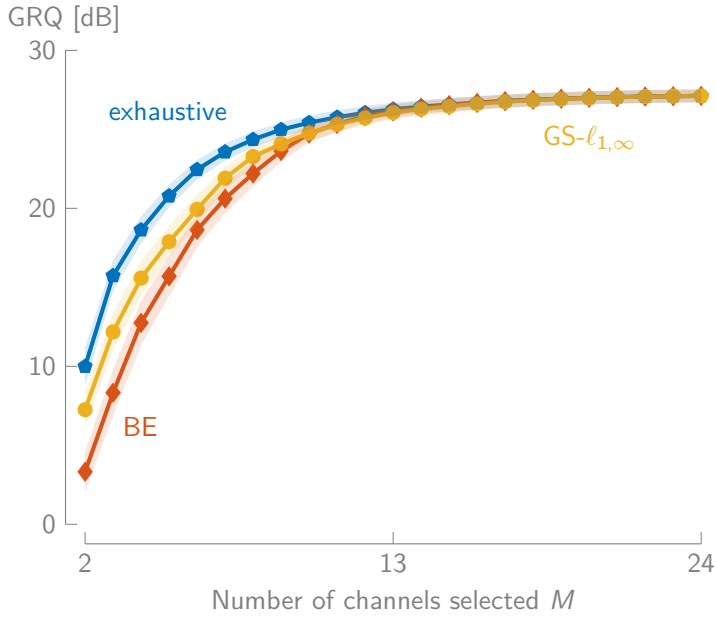


Figure 3.4.4: While the BE and GS-$\ell_{1,\infty}$ method performs on par for large M, the GS-$\ell_{1,\infty}$ method starts to outperform the BE method for smaller M in the ill-conditioned $\mathbf{R}_2$ covariance matrix case (mean $\pm$ standard error on the mean).

Figure 3.4.5 shows the differences in GRQ across all runs and M between 2 and 12 between the BE/GS-$\ell_{1,\infty}$ method and the exhaustive solution when $2 \leq N_2 \leq 12$. The BE method has a heavier tail with more outlying negative differences with the exhaustive solution than the GS-$\ell_{1,\infty}$ method. Of all runs with $2 \leq N_2 \leq 12$, there is a fail rate of 11.52% for the BE method, while this is only 0.91% for the GS-$\ell_{1,\infty}$ method. To summarize, when there is an ill-conditioned covariance matrix $\mathbf{R}_2$ in the denominator of the GRQ, the GS-$\ell_{1,\infty}$ method is more robust than the BE method.

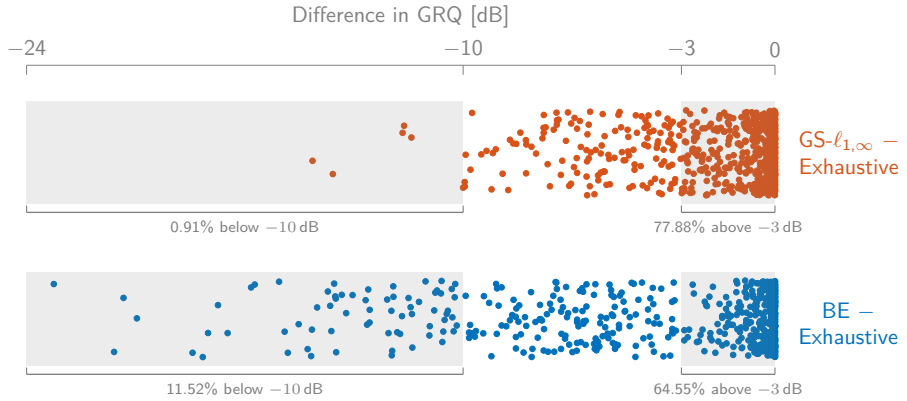


Figure 3.4.5: The BE method shows more outlying negative differences in GRQ (across all runs and M between 2 and 12) with the exhaustive solution than the GS-$\ell_{1,\infty}$ method when the covariance matrix in the denominator of the GRQ is ill-conditioned.

The failure cases of the BE method can be reduced by preventing covariance matrix $\mathbf{R}_2$ from being ill-conditioned. This can be done by performing the BE method on a regularized covariance matrix $\tilde{\mathbf{R}}_2$. The regularization should increase the condition number of the covariance matrix $\tilde{\mathbf{R}}_2$. The solution of the BE method on the regularized problem should provide a more robust estimate of the optimal solution. Several regularization techniques can be used including diagonal loading or the subspace projection presented in the previous chapter (sub-section 2.2.2). Diagonal loading has been widely used to improve robustness of GEVD problems, such as beamforming, in presence of noise [125]. Diagonal loading is applied by adding an identity matrix ($\mathbf{I}$) to the covariance matrix of interest:

$$\tilde{\mathbf{R}}_2 = \mathbf{R}_2 + \lambda \mathbf{I}$$

The relative power of the identity matrix ($\mathbf{I}$) in relation to the covariance matrix ($\mathbf{R}_2$) is a parameter (λ) which must be carefully selected. One estimator for an

optimal choice of this parameter is described as the Ledoit-Wolf estimator. The estimator is shown to minimize the mean squared error between an estimated covariance matrix acquired from a finite set of observations and the real covariance matrix [126].

To evaluate the effect of regularization, the following simulation is conducted. An artificially rank-deficient problem is created. The simulation setup in section 3.4.2 is used to generate a grid of 9 sensors ($L = 1, K = 1$). 9 additional sensors are constructed as random linear combinations of the first 9 sensors. They are added to create a pool of 18 sensors (half of which are redundant). 400 Monte Carlo runs are executed. The results of the simulation are shown in figure 3.4.6. It shows the BE method with Ledoit-Wolf regularization outperforms standard BE for any number of sensors. A comparison to the exhaustive solution is used to compare the failure cases of the standard BE with the Ledoit-Wold regularized BE. In these simulations with 9 independent sensors and 9 redundant sensors, the standard BE fails (defined here as $> 10\,\text{dB}$ difference with the optimal exhaustive solution) in 35% of cases. This failure rate decreases to 19% with the regularized BE. This represents a reduction of 47% of the failure cases. This shows regularization of the covariance matrix ($\mathbf{R}_2$) can improve the robustness of the BE method.

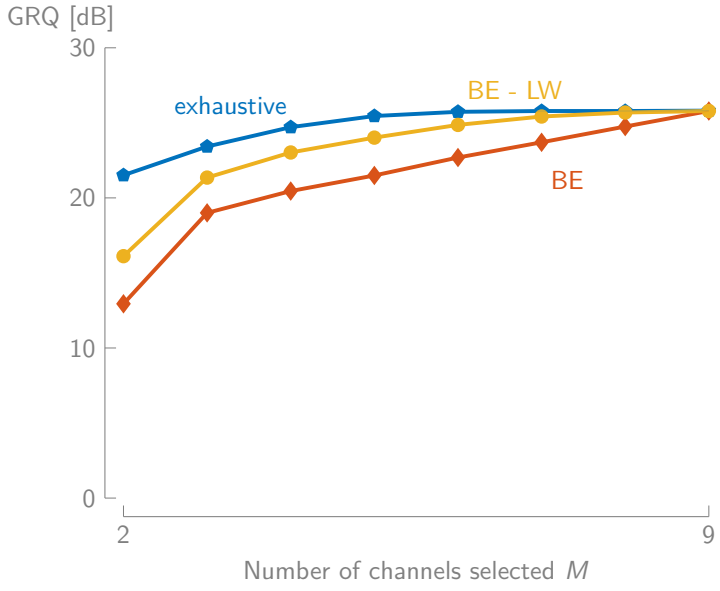


Figure 3.4.6: The BE with Ledoit-Wolf regularization (BE-LW) outperforms standard BE for any number of sensors.

3.5 Example of sensor selection on real-world data

The benchmark study in Section 3.4 was performed on simulated data, which allowed us to generate a large number of simulations that are generic enough to apply to many sensor selection problems that arise in different signal processing domains. In this section, we show an example of sensor selection in the context of mobile epileptic seizure monitoring. More specifically, the task requires to design a spatiotemporal filter that amplifies multi-channel EEG data during seizures while maximally attenuating peak interferers [104]. The solution is found through max-SNR filtering and can be solved using the GEVD framework described in this chapter. More information about the context, problem, and data can be found in chapter 2.

In the following example, we investigate the effect of the reduction of EEG channels on subject three of the study in chapter 2, aiming to design a mobile EEG setup. The data contains 16 channels (i.e., $C = 16$). Five time lags are used per channel (i.e., $L = 5$), while two output filters are computed (i.e., $K = 2$).

Figure 3.5.1 shows the GRQ as a function of the number of selected channels for the GS-$\ell_{1,\infty}$, the FS and BE methods, which performed best in the benchmark study (see Section 3.4.3 to 3.4.5). The results are in line with the benchmark study. Both the GS-$\ell_{1,\infty}$ and BE methods obtain similar GRQ across the whole range of selected channels, but the GS-$\ell_{1,\infty}$ and FS methods outperform the greedy BE method for a low number of channels M. This confirms that the developed method performs as expected, also on sensor selection problems on real-world data.

The example of channel selection on a standard EEG cap ($C = 16$) allows using the GS-$\ell_{1,\infty}$ method, the FS or the BE method with confidence on the selected channels with regards to the optimal exhaustive selection. When the number of available channels increases as is found in high density EEG caps with 256 channels, the computational complexity of the different methods becomes a determinant factor when considering the channel selection method. The computational time of the different methods on an Intel quad-Core i7-4790 (3.60 GHz) for the selection of 6 channels in a problem with 64 channels ($C = 64, L = 1, K = 1$) is shown in table 3.5.1. The FS method is 20 times faster than the BE method to select 6 channels out of a pool of 64 ($L = 1, K = 1$). Both these methods obtain a result in less than one second while the GS-$\ell_{1,\infty}$ takes more than 15 minutes. The computation of all these methods increases with the number of channels (C), the numer of time lags (L) and filter outputs (K) such that the GS-$\ell_{1,\infty}$ is impractical to use in problems with many variables.

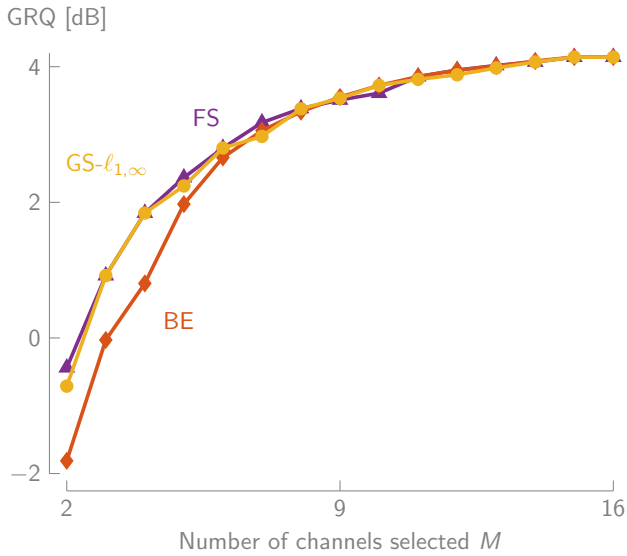


Figure 3.5.1: The GS-$\ell_{1,\infty}$ and the FS methods outperform the BE method for a low number of channels also on an example with real-world data collected on a patient with epilepsy ($C = 16, L = 5, K = 2$).

	FS	**BE**	GS-$\ell_{1,\infty}$
Compute time [s]	0.02	0.4	924

Table 3.5.1: The FS method is 20 times faster than the BE method to select 6 channels out of a pool of 64 ($L = 1, K = 1$). Both these methods obtain a result in less than one second while the GS-$\ell_{1,\infty}$ takes more than 15 minutes.

3.6 Conclusion

In this chapter, we proposed a group-sparse variable selection method using the $\ell_{1,\infty}$-norm for GRQ optimization and GEVD problems applied in the context of sensor selection. This group-sparsity does not only allow to extend spatial to spatio-temporal filtering but also to take multiple filters (eigenvectors) into account and thus extend MISO to MIMO filtering. The latter is essential in various other applications, such as selecting sensors across different filterbands in CSP applications [109], [110].

We have extensively compared the proposed GS-$\ell_{1,\infty}$ method with various other sensor selection methods (greedy, optimization-based, ...). Remarkably, the simple greedy BE method outperformed all methods from the state of the art, except the proposed GS-$\ell_{1,\infty}$ method. While the heuristic BE method is computationally more efficient, it performs worse than the GS-$\ell_{1,\infty}$ method for smaller numbers of selected sensors, and with a higher probability to completely fail. We have shown that one specific fail case of the BE method is when the covariance matrix in the denominator of the GRQ is ill-conditioned. We also showed the FS method performs in line with the GS-$\ell_{1,\infty}$ method when the number of sensors to select is small.

As the BE method is less robust than the proposed GS-$\ell_{1,\infty}$ method, the latter is the preferred choice when performing variable selection, in particular if the number of desired variables is small compared to the total number of variables. In cases where the total number of variables is large such that the GS-$\ell_{1,\infty}$ method becomes computationally impractical, the FS or regularized BE are the preferred methods for respectively low and high number of desired variables compared to the total number of variables.

In the next chapter, we will be investigating the effects of miniaturization on a setup with 256 EEG channels. From this setup we will be selecting one to ten channels. For this problem which requires to select a low number of variables from a large pool, this chapter showed the FS method is the best available method.

CHAPTER 4

Sensor selection and miniaturization limits for detection of interictal epileptiform discharges with wearable EEG

This chapter is based on J. Dan, M.T. Foged, B. Vandendriessche, W. Van Paesschen and A. Bertrand, "Sensor selection and miniaturization limits for detection of interictal epileptiform discharges with wearable EEG", Under review, 2022.

Sensor selection and miniaturization limits for detection of interictal epileptiform discharges with wearable EEG

Abstract | *Objective* The goal of this chapter is to investigate the limits of EEG sensor miniaturization in a set-up consisting of multiple galvanically isolated EEG units to record interictal epileptiform discharges (IED), referred to as 'spikes', in people with epilepsy. *Approach* A dataset of high-density EEG recordings (257 channels) was used to emulate local EEG sensor units with short inter-electrode distances. A computationally efficient sensor selection and interictal spike detection algorithm was developed and used to assess the influence of the inter-electrode distance and the number of such EEG units on spike detection performance. Signal-to-noise ratio, correlation with a clinical-grade IED detector and Cohen's kappa coefficient of agreement were used to quantify performance. Bayesian statistics were used to confirm the statistical significance of the observed results. *Main Results* We found that EEG recording equipment should be specifically designed to measure the small signal power at short inter-electrode distance by providing an input referred noise < 300 nV. We also found that an inter-electrode distance of minimum 5 cm between electrodes in a setup with a minimum of two EEG units is required to obtain near equivalent performance in interictal spike detection to standard EEG. *Significance* These findings provide design guidelines for miniaturizing EEG systems for long term ambulatory monitoring of interictal spikes in epilepsy patients.

4.1 Introduction

Epilepsy is one of the most common severely disabling brain conditions, affecting over 46 million people worldwide [57]. The disorder is characterized by pathological electrical discharges of neurons, which can be recorded non-invasively using scalp EEG as isolated spikes or spike-wave complexes. The latter occur during clinical epileptic seizures, but also between seizures (i.e. IED or 'spikes'), without any concomitant clinical manifestations. EEG is routinely used in the clinic to document such discharges in order to contribute to diagnosis, follow up, and adaptation of treatment for people with epilepsy. Clinical EEG recordings are typically carried out in a hospital (outpatient or inpatient) setting in a resting condition over a relatively short period of time. The American Clinical Neurophysiology Society recommends, that this lasts at least 20 minutes [15]. Therefore, routine EEG often provides only a snapshot of the disorder experienced by a person with epilepsy. Short duration EEG recordings cannot capture long-term patterns of the disorder such as seizure frequency, timing or dynamic cycles [127]. Long-term monitoring of interictal spikes has been shown to be an effective predictor of seizure timing [128] as well as treatment outcome [129]. In order to obtain long-term EEG in realistic conditions, miniaturized EEG devices that can be worn by people in their everyday life would be preferable.

Previous work has demonstrated that EEG recording using several dozens of electrodes, such as those used in clinical routine, are not required to effectively monitor discharge features in people with a diagnosis of epilepsy and that a low number of electrodes can be sufficiently informative [104], [130], [131]. Furthermore, wearable EEG devices for everyday life monitoring are in active development [35], [132]–[135]. Efforts to miniaturize EEG devices have mainly focused either on the electronics of the recording device or on the electrode placement setup. Some examples include the work of Zibrandtsen et al. [135] who made custom fitted in-ear EEG to monitor people with epilepsy. They analyzed intra-ear EEG channels and inter-ear EEG channels and investigated the feasibility of annotating epileptic seizures based on in-ear EEG channels alone. They found ear-EEG can detect temporal lobe seizures and generalised seizures as well as interictal spikes. Swinnen et al. [35] used Byteflies Sensor Dot, a commercial miniature EEG sensor with two channels, to detect typical absence seizures in adults and children. They found epileptologists were able to reliably detect typical absence seizures using this miniature EEG sensor.

One way of further minimizing the EEG setup is by reducing the number of wires. This can be done by using several EEG amplifiers that are each connected to a single pair of electrodes. These isolated EEG sensor units measure a local

bipolar channel, each with a built-in local amplifier. Such a set-up with multiple miniaturized and galvanically isolated EEG sensor units eliminates the need for multiple long wires that run across the scalp, and is sometimes referred to as a wireless EEG sensor network [44], [45], [47]. The bulkiness of such an EEG sensor network is determined by the number of sensor units and the size of these units. The latter is mostly dependent on the distance between the electrodes of each such sensor unit.

Obviously, reducing the number of recording electrodes increases the importance of electrode placement as it is expected to influence the detection of epileptiform discharges, even for generalised discharges [130]. As a corollary, it is critical to individualize the placement of reduced numbers of electrodes to the specifics of a patient discharge pattern as compared to routine hospital-based recordings.

To approach this electrode placement problem, and to analyze miniaturization effects, a setup with multiple sensor units with short bipolar inter-electrode distances can be emulated and evaluated using information collected through high density electroencephalograph (HD-EEG) recordings. In a 257-electrode HD-EEG head cap, electrodes are situated approximately 2 centimeters apart, thus allowing to emulate short inter-electrode distances. This type of approach has been previously proposed in the context of auditory attention decoding, to identify and quantify the lower bound on miniaturization for EEG-based decoding of neural responses to speech. Mundanad Narayanan et al. showed that for inter-electrode distances equal or greater than 3 cm, the decoding performance was not significantly worse compared to that achieved with long distance channels referenced to the `Cz` electrode [47]. Furthermore, they showed a rapid decline in decoding performance for inter-electrode distances smaller than 3 cm. The study was also performed on an HD-EEG dataset where short distance channels were constructed through re-referencing. Studies on the effect of inter-electrode distance had already been reported in the 1980. Authors showed that the amplitude of the EEG follows an exponential relationship with inter-electrode distance [136]. However, in the context of spike detection, it is not necessarily the decrease in amplitude that impacts the detection performance, but (also) the changes in signal-to-noise (SNR) as the noise levels will also change with a decrease in inter-electrode distance.

In this chapter, we investigate the limits of miniaturization for interictal spike detection using a network of galvanically isolated EEG units by emulating such a set-up via HD-EEG recorded on people with epilepsy. We present an automated algorithm pipeline for the selection of a small number of EEG channels for mobile EEG ambulatory monitoring and for the automatic detection of interictal epileptiform discharges. The effects of miniaturization, both in terms of short inter-electrode distance and in terms of a small number of EEG units, were assessed through comparison with spike detection on standard 32 channels

of 'long-distance' EEG in which all electrodes are wired to a central reference electrode.

Processing of HD-EEG is a computationally heavy operation due to the high number of electrodes resulting in high-dimensional data. In order to study the influence of the number of EEG sensor units and the inter-electrode distance, efficient methods to process these data have to be developed. First a computationally efficient spike detector should be used. Spike events are stereotypical in that spatial and temporal signature of different events are very similar (within the same patient). Algorithms that leverage this build an average spike template that is then used as a matched filter for detecting spikes [137]–[140]. This type of algorithm is computationally efficient due to the low number of operations required to classify a new epoch of data. In this chapter, we use an algorithm that leverages the spatio-temporal signature of spike events by constructing a filter that enhances spikes, but also suppresses non-spike EEG activity and artefacts by exploiting the spatio-temporal second order statistics of the latter components. In this way, the filter maximizes the SNR (i.e. spike to noise ratio) to obtain a more accurate detection than a straightforward matched filter. This filter design paradigm is known as max-SNR filtering [82]. We then select a small number of EEG units using a channel selection procedure to obtain a reduced channel set from a large set of candidate channels. This must also be computationally efficient and should make a selection that is optimized for the spike detection task. In chapter 3, we compared different channel selection methods for group-sparse generalised eigenvalue decomposition (GEVD) problems, which is a general class of problems that contains the max-SNR filter as a special case. The group-sparsity appears in the spatio-temporal filter, where selecting one EEG channel corresponds to also selecting all the time lags associated with that channel (representing a group of optimization variables). In the chapter, we found that for a low number of channels to be selected, a forward greedy selection was the most efficient, while being competitive with more expensive optimization methods.

Manual annotation of spike events is a time-consuming activity that requires the expertise of experienced clinical neurophysiologists [141]. It is a difficult task that must meet the six criteria suggested by the International Federation of Clinical Neurophysiology [19]. These criteria define an IED as (1) di- or tri-phasic waves with sharp or spiky morphology (i.e. pointed peak); (2) different wave duration than the ongoing background activity: either shorter or longer; (3) asymmetry of the waveform: a sharply rising ascending phase and a more slowly decaying descending phase, or vice versa; (4) the transient is followed by an associated slow after-wave; (5) the background activity surrounding IED is disrupted by the presence of the IED; and (6) distribution of the negative and positive potentials on the scalp suggests a source of the signal in the brain,

corresponding to a radial, oblique, or tangential orientation of the source. While these criteria provide a basis for labeling a spike, inter-rater agreement is only fair, with reported Cohen's kappa coefficient of 0.49 for inter-rater agreement for single spike identification [142], 0.63 for the occurrence of one or more spikes during a 10-second epoch, and 0.69 for a 30-minute epoch [143]. Because of the time required, it is rare for human experts to fully annotate an EEG recording. For diagnostic purposes, the annotation of several clear spike events that allow confirmation of the diagnosis of epilepsy and localization of the seizure onset zone are often sufficient. Human annotation can be facilitated by a spike detection software. Only a few software packages are available on the market with certification for use in clinical practice. Currently, Persyst is the spike detection software with the highest agreement to human experts [143]. It is a fully automated spike detector based on a deep neural network.

The outline of the chapter is as follows. We present the dataset, along with the channel selection and spike detection methodology in section 4.2. Next, in section 4.3, we present our results on the effect of inter-electrode distance and number of miniaturized EEG units on the ability to identify spikes. They are discussed and put in perspective with the current corpus of research in section 4.4. Conclusions and implications of the work are highlighted in section 4.5.

4.2 Methods

This section describes the material and methods that were used to conduct the analysis on the limits of EEG sensor miniaturization to record interictal spikes. We first describe the EEG data set in subsection 4.2.1, after which we describe the different steps of our signal processing methodology in the remaining subsections. The signal processing pipeline consists of the following steps. First, a base set of annotations for spikes are generated using the automated spike detector provided with the Persyst software. Then, a pool of bipolar channels with a fixed inter-electrode distance is generated by re-referencing the high-density EEG channels.A spike detector based on a max-SNR filtering criterion is developed. A channel selection algorithm which optimizes this max-SNR criterion is used to select the N best channels. The spike detector is then trained on these N channels. A detailed description of each of these steps is presented in the next subsections along with the performance metrics used to evaluate miniaturization limits.

4.2.1 Recordings

The recordings used in this study were originally obtained between 2015 and 2018 in the context of pre-surgical evaluation of patients with drug-resistant focal epilepsy referred to a tertiary care hospital (Rigshospitalet, Copenhagen, Denmark). HD-EEG recordings were performed on 40 subjects over 120 minutes with 257 electrodes with exchangeable sponge electrodes using a EGI Gedodesic Sensor Net HydroCel GSN 130 amplifier and cap sampled at 1 KHz.

4.2.2 Spike annotations

In order to automate the annotation of IEDs (from hereon referred to as 'spikes'), the recordings were analyzed using Persyst version 14.D software (Persyst Development Corporation, Solana Beach, CA, USA) to annotate the spikes. Persyst is a clinical grade EEG analysis software that specializes in epilepsy diagnosis. It contains a fully automated spike detector based on a deep neural network. To obtain the spike annotations, the software's low-pass and high-pass values were left at their default values (0.16 Hz - 70 Hz). The notch filter was set to 50 Hz. The detection was set to the low sensitivity setting (Persyst score > 0.9), which has high specificity [143]. Spikes occurring during periods contaminated by strong artifacts (peak amplitude $> 200\ \mu V$ or 1 second root-mean-square $> 100\ \mu V$, which are mostly associated with tapping and movement artifacts) and recordings with a low number of spikes such that cross-validation could not be performed (less than 10 spike events) were excluded from the analysis.

A total of 40 subjects were processed by Persyst for spike annotations. Fifteen subjects had ten or more spike events meeting the inclusion criteria and were included in the analysis. A total of 4088 spikes were detected by Persyst. The median number of spikes per subject was 58 (range 27-1190) The number of spike events detected by Persyst are given per subject in table 4.2.1.

	# spikes
subject 1	1190
subject 2	1016
subject 3	439
subject 4	401
subject 5	375
subject 6	231
subject 7	111
subject 8	58
subject 9	47
subject 10	45
subject 11	43
subject 12	40
subject 13	38
subject 14	27
subject 15	27
Total	4088

Table 4.2.1: Number of interictal epileptiform discharges detected by Persyst 14 at the low sensitivity setting.

4.2.3 Miniature EEG emulation

The miniaturization analysis was performed by extracting a pool of candidate channels (each representing a separate galvanically isolated EEG sensor unit), with desired inter-electrode distances of either 2, 3.5, 5, 6.5 or 8 cm between electrode pairs (each distance leads to a different pool of candidates which are then standardized across scenarios as explained below). We refer to the inter-electrode distance with the value d. To obtain this pool of bipolar channels, we implemented the following procedure. First, the inter-electrode distance between all possible electrode pairs was computed. The latitude and longitude of each electrode in the EGI Gedodesic Sensor Net HydroCel GSN 130 headset is given by the manufacturer. The distance was computed as the haversine distance with the head modeled as a sphere with a circumference of 57 cm [144]. The haversine distance or great-circle distance is the shortest distance between two points on the surface of a sphere, measured along the surface of the sphere [145]. In the context of EEG sensor units, this distance corresponds to the length of the wire between two electrodes. Figure 4.2.1 shows the number of electrode pairs (channels) as a function of the inter-electrode distance. The electrode pairs at a distance $d \pm 0.25$ cm were retained. Since each d results in a different number of candidate channels, a channel selection procedure was then

applied to each of the pools in order to obtain an equal number of candidate channels independently of inter-electrode distance d. This selection eliminates redundant channels to obtain a pool of candidate channels that uniformly cover the scalp. Channels were considered redundant with respect to each other if they were geometrically close by and in the same orientation as they would record very similar signals. However, electrode pairs with close midpoints but orthogonal orientation would record relevant information as they record dipoles with different orientations. To this end, we use a proximity metric (p), which combines the distance (d) and the orientation between two channels, based on

$$d = \text{haversine\_distance}(\mathbf{m_1}, \mathbf{m_2}) \tag{4.1}$$

$$p = e^{(-\frac{2}{3} * d)} * |\mathbf{v_1} \cdot \mathbf{v_2}| \tag{4.2}$$

where $\mathbf{m_1}, \mathbf{m_2}$ are the midpoints of each channel (i.e. the center point of the great circle line between the two electrodes that define the channel), $\mathbf{v_1}, \mathbf{v_2}$ are the unit orientation vectors of each channel (defining the orientation of the line between the two electrodes that define the channel), and d is the harvesine distance between the midpoints of both channels. The different symbols used in the computation of the proximity metric are illustrated in figure 4.2.2. The proximity metric p in equation 4.2 is defined such that it is high if the two channels are close to each other (small d), *and* if their orientation is similar (large $|\mathbf{v_1} \cdot \mathbf{v_2}|$). Indeed, if either the angle between $\mathbf{v_1}$ and $\mathbf{v_2}$ is near-orthogonal or the distance d is large, then p will be small. The local density of a channel was computed as the sum of the proximity p of a channel to all other channels. To obtain a set of 256 channels, the channel with the highest proximity to all other channels was iteratively removed until a set of 256 channels was obtained in each candidate pool.

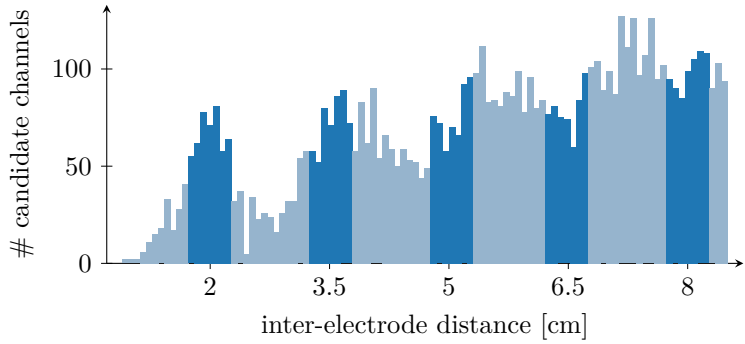


Figure 4.2.1: Histogram of the inter-electrode distance between all electrodes in a EGI Geodesic Sensor Net HydroCel GSN 130 headset with 257 electrodes. Distances of 2, 3.5, 5, 6.5 and 8 cm ±0.25 cm, highlighted in darker blue, are investigated in this study.

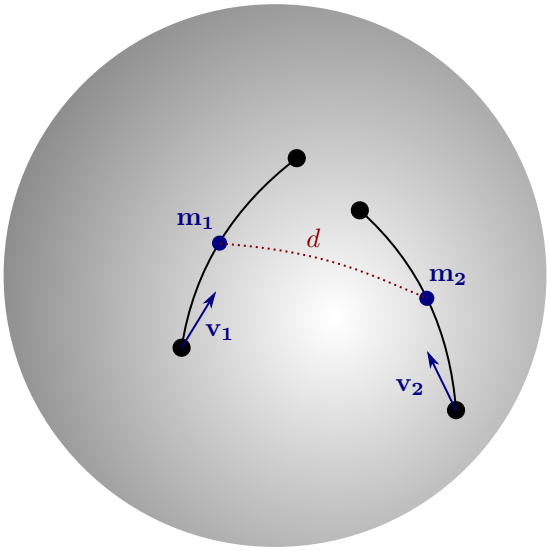


Figure 4.2.2: Illustration of the different symbols used in the computation of the proximity between two pairs of electrodes.

4.2.4 Spike detection algorithm

Template matching has been demonstrated to successfully detect neural spiking activity both on intracranial and scalp EEG recordings [137]–[140]. In this chapter, we use a filter design framework that is akin to such a template matching filter, with the additional benefit that it takes the noise statistics into account (in this case any non-spike EEG activity) in order to maximize the SNR for the detection. The max-SNR filtering algorithm we use here is a variation of an algorithm we previously developed and validated for the detection of epileptic seizures [104]. It is a pre-trained multi-channel filter, implemented as a filter-and-sum pipeline. The algorithm requires examples of spikes and noise (i.e. non-spike EEG), which are obtained automatically via Persyst executed at a high specificity level (see below).

The raw signals were first downsampled from 1 KHz to 100 Hz, then bandpass filtered between 3 Hz and 30 Hz using a zero-phase, non-causal bandpass filter. The EEG signal in channel k at sample time index t is modeled as

$$x_k(t) = s_k(t) + n_k(t)$$

where $s_k(t)$ corresponds the signal component that contains all the spike events and $n_k(t)$ corresponds to background EEG, which in our case is considered to be noise. The N-channel EEG signal is denoted as $\mathbf{x}(t) \in \mathbb{R}^N$ with $\mathbf{x}(t) = [x_1(t) \dots x_N(t)]$. The aim is to produce a filter $\mathbf{w} \in \mathbb{R}^N$ that filters and combines the N channels of EEG into a single-channel output signal $o(t)$ in which the background EEG is maximally suppressed, while preserving the spikes. This output channel is obtained through the linear combination:

$$o(t) = \mathbf{w}^T \mathbf{x}(t) \tag{4.3}$$

From a signal processing viewpoint, $\mathbf{w}$ acts as a spatial filter that linearly combines different EEG channels at different positions on the scalp. The filter $\mathbf{w}$ is optimized in a data-driven fashion to maximize the SNR of $o(t)$ over a training set, i.e. solving

$$\max_{\mathbf{w}} \frac{E\{(\mathbf{w}^T \mathbf{s}(t))^2\}}{E\{(\mathbf{w}^T (\mathbf{n}(t))^2\}} = \max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{R_s} \mathbf{w}}{\mathbf{w}^T \mathbf{R_n} \mathbf{w}} \tag{4.4}$$

where $E\{.\}$ denotes the expected value operator, $\mathbf{R_s} = E\{\mathbf{s}(t)\mathbf{s}(t)^T\}$ and $\mathbf{R_n} = E\{\mathbf{n}(t)\mathbf{n}(t)^T\}$ denote the spike and noise covariance matrices, respectively, and $\mathbf{s}(t)$ and $\mathbf{n}(t)$ are defined similarly to $\mathbf{x}(t)$, i.e., $\mathbf{x}(t) = \mathbf{s}(t) + \mathbf{n}(t)$. In subsection 4.2.4, we will explain how these covariance matrices $\mathbf{R_s}$ and $\mathbf{R_n}$ are estimated from the data. It can be shown [82] that the solution of the maximization problem defined in equation 4.4 is the eigenvector corresponding

to the largest eigenvalue of the matrix $\mathbf{R_n^{-1}R_s}$. This is equivalent to solving a GEVD problem based on the matrix pencil $(\mathbf{R_s}, \mathbf{R_n})$ [82].

The filter described above is a spatial filter. It can be expanded to a causal spatio-temporal filter by creating a buffer of L samples for each channel and stacking all buffered (time-lagged) samples in a single vector $\tilde{\mathbf{x}}(t) = col\{\tilde{\mathbf{x}}_\mathbf{1}(t), \dots, \tilde{\mathbf{x}}_\mathbf{N}(t)\}$ where $\tilde{\mathbf{x}}_\mathbf{k}(t) = [x_k(t),\ x_k(t-1), \dots, x_k(t-L+1)]^T$ and $col\{.\}$ denotes a columnwise stacking. The output signal $o(t)$ is given by equation 4.3 where $\mathbf{w}$ is replaced by $\tilde{\mathbf{w}} \in \mathbb{R}^{LN}$ and $\mathbf{x}(t)$ is replaced by $\tilde{\mathbf{x}}(t) \in \mathbb{R}^{LN}$. This then corresponds to a filter-and-sum operation, where each channel is filtered with a channel-specific finite impulse response filter of length L, followed by a summation across channels. For spike detection the buffer length is set to 200 ms ($L = 20$ at 100 Hz sampling), corresponding to the full duration of a spike discharge.

In this spatio-temporal extension, the covariance matrices in equation 4.4 are replaced with their spatio-temporal generalizations, i.e., $\mathbf{R_s} = E_s\{\tilde{\mathbf{s}}(t)\tilde{\mathbf{s}}(t)^T\}$ and $\mathbf{R_n} = E_i\{(\tilde{\mathbf{n}}(t))(\tilde{\mathbf{n}}(t))^T\}$. In the remainder of the chapter, we always assume the spatio-temporal extension unless otherwise specified, and omit the $\tilde{}$ for notational convenience.

Covariance matrix estimation

In order to compute the optimal filter $\mathbf{w}$, we need to estimate the two covariance matrices $\mathbf{R_s}$ and $\mathbf{R_n}$ on a training data set.

In our implementation of the algorithm, the spike covariance matrix $\mathbf{R_s}$ is estimated based on a training set $\mathcal{S}$ with a few example [1] spike waveforms of the patient under test. To this end, we used the following estimator [2]

$$\mathbf{R_s} \approx \frac{1}{L|\mathcal{S}|} \sum_{\tau \in \mathcal{S}} f(\tau) \sum_{t=\tau}^{\tau+L} \mathbf{x}(t)\mathbf{x}(t)^T \tag{4.5}$$

$$f(\tau) = g\left(\sum_{t=\tau}^{\tau+L} ||\mathbf{x}(t)^2||_1 \left(\frac{\sum_{\nu \in \mathcal{S}} \sum_{t=\nu}^{\nu+L} ||\mathbf{x}(t)^2||_1}{|\mathcal{S}|} \right)^{-1} \right) \tag{4.6}$$

[1] These examples can be obtained, e.g., from a manual annotation on part of the data, or via automatic annotation with clinical-grade software such as Persyst in high-specificity modus to extract a few examples with a minimal number of false positives.

[2] Note that, since the noise floor is always present, the resulting covariance matrix is actually an estimate for $E\{(\mathbf{s}(t) + \mathbf{n}(t))(\mathbf{s}(t) + \mathbf{n}(t))^T\} = \mathbf{R_s} + \mathbf{R_n}$. However, since replacing $\mathbf{R_s}$ with $(\mathbf{R_s} + \mathbf{R_n})$ in (4.4) leads to the equivalent optimization problem $\max_\mathbf{w} 1 + \frac{\mathbf{w}^T\mathbf{R_s}\mathbf{w}}{\mathbf{w}^T\mathbf{R_n}\mathbf{w}}$, we obtain the same optimal filter.

$$g(x) = \begin{cases} 1 & \text{if } x < 1 \\ \frac{1}{x} & \text{if } x >= 1 \end{cases} \tag{4.7}$$

where $\mathcal{S}$ is the set of time indices corresponding to the first time sample of each 'spike' segment marked by the Persyst spike detector. Each spike has a duration of 200 ms. $f(\tau)$ is a normalization of the power of a spike event with respect to the average power of all spike events for events with above-average power. It is used to reduce the influence of outlier spikes with high power (i.e., above-average). The noise covariance matrix $\mathbf{R_n}$ was computed similarly without the normalization factor as

$$\mathbf{R_n} \approx \frac{1}{|\mathcal{N}|} \sum_{t \in \mathcal{N}} \mathbf{x}(t)\mathbf{x}(t)^T$$

where $\mathcal{N}$ is a set of training samples that do not contain a spike event. Since most of the data does not contain spikes (as IEDs are sparse events), it is not a problem if some spikes leak into $\mathbf{R_n}$, as they will not have a large impact in the overall sum.

Given the cross-validation scheme, $\mathbf{R_s}$ is evaluated on 20 to 892 spikes depending on the test subject. $\mathbf{R_s}$ is always evaluated on 1h30 of data.

Regularization

The inclusion of time lags substantially increases the dimension of the covariance matrices, possibly making them ill-conditioned due to redundancy in the entries of $\mathbf{x}(t)$. For this reason a regularization scheme is required. Wouters et al. [83] proposed an effective regularization scheme for a template-matching-filter that optimizes a similar cost function as in equation 4.4, but where the numerator consists of a single template instead of a covariance matrix. The regularization is obtained by projecting the data on a subspace containing the main principal components of the denominator covariance matrix along with the template itself (to represent the target signal). In this work we adapted the method to our problem formulation. The data were projected on a subspace defined as the span of the principal components of both the noise $\mathbf{R_n}$ and spikes $\mathbf{R_s}$ covariance matrices. The principal components with the largest eigenvalues of $\mathbf{R_n}$ and accounting for 85% of the variance in the noise segments were retained. The principal components with the largest eigenvalues of $\mathbf{R_s}$ accounting for 95% of the variance in the spike segments were also retained. These two sets of principal component vectors were then combined and orthogonalized by placing them in the columns of a new matrix $\mathbf{M}$ on which a singular value decomposition is applied to find an orthogonal basis for the subspace. Let $\mathbf{U} = [\mathbf{u}_1 \dots \mathbf{u}_K]$ denote the matrix, the columns of which consist of the K left singular vectors

corresponding to the largest singular values of $\mathbf{M}$, where the cut-off K is chosen such that the cumulated sum of these singular values is at least 99% of the sum of all singular values. The matrix $\mathbf{U}$ is then used as a compression matrix on the data, i.e.

$$\mathbf{x_c}(t) = \mathbf{U}^T\mathbf{x}(t)$$

where $\mathbf{x_c}(t) \in \mathbb{R}^K$. The optimal compressed filter $\mathbf{w_c}$ is then given as the eigenvector of $(\mathbf{R_{n,c}^{-1}})\mathbf{R_{s,c}}$, where $\mathbf{R_{n,c}} = \mathbf{U}^T\mathbf{R_n}\mathbf{U}$ and $\mathbf{R_{s,c}} = \mathbf{U}^T\mathbf{R_s}\mathbf{U}$ correspond to the compressed matrices. The filter output is then defined as $o(t) = \mathbf{w_c}^T\mathbf{x_c}(t)$ which is equivalent to an uncompressed filtering $o(t) = \mathbf{w}^T\mathbf{x}(t)$ with $\mathbf{w} = \mathbf{U}\mathbf{w_c}$. It is noted that this compression is only applied with the purpose of regularization during training to obtain a better (uncompressed) filter $\mathbf{w}$. During operation of the algorithm (at test time), we always apply the full filter $\mathbf{w}$ on the uncompressed data, as the compression with $\mathbf{U}$ requires more computations than the filtering with $\mathbf{w}$.

Classification

The classification of an epoch as containing spikes was performed twice per second (i.e. at 0.5 Hz). As the filter is trained to suppress non-spike EEG and enhance spikes, we can perform detection based on the output power or root mean square (RMS) value of the filter. Therefore, as features for the classifier, we used the moving average RMS value of the single-channel filter output $o(t)$ averaged over 200 ms (i.e. the duration of a spike) and the moving average RMS averaged across the different channels of the EEG data $\mathbf{x}(t)$ and averaged over one second (the latter serves as a baseline reference). The downsampling to 0.5 Hz was done with a bucket algorithm [146]. This algorithm preserves the amplitude of short duration local extrema such as spikes that would otherwise be attenuated by a standard low-pass and decimation algorithm. These two RMS features were log transformed before being fitted by a linear discriminant analysis (LDA) classifier, which was trained based on the spike annotations in the training set.

4.2.5 Greedy Forward Channel selection

To reduce the number of channels from 256 candidates, a channel selection procedure is required. As we are assessing the performance of a spike detection task based on a max-SNR framework, which is equivalent to solving a GEVD problem [82], the channel selection task corresponds to a (group)-sparse GEVD problem. In chapter 3 the performance of various group-sparse variable selection methods for GEVD problems has been investigated. We showed that greedy

forward selection performed competitively with other state-of-the-art methods at a significantly lower computational cost. In chapter 3 we found the proposed GS-$\ell_{1,\infty}$ performed best for a low number of channels to select from a large pool of candidates. However, that method is prohibitively slow for the large problem at hand with 256 channels and several time lags ($L = 4$). The second-best method when few channels need to be selected from a large pool of candidates is the forward selection. The forward selection method starts from an empty set of channels $\mathcal{C}$ and sequentially adds the channel that maximally increases the objective in equation 4.4 when added to the current set $\mathcal{C}$. New channels are added until $\mathcal{C}$ contains M channels, where M is a pre-defined number of required channels. Some examples of selected channels are shown in figure 4.2.3.

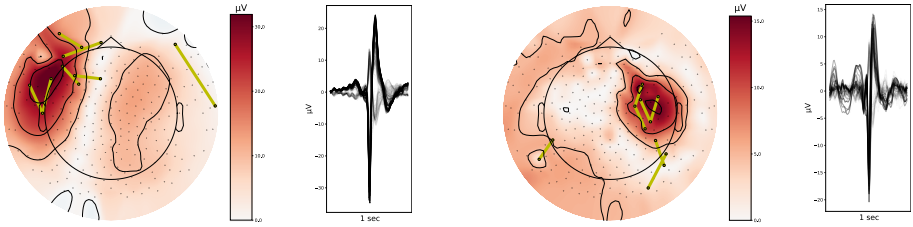


Figure 4.2.3: Example of seven channels selected in two example subjects. The inter-electrode distance of the channels is set to 5 cm. The topoplot shows the magnitude of the average spike of Persyst at the peak of the spike (t=0) in a common average montage. The 1 second timecourse of the average spike is shown with all channels superimposed. The figure illustrates the automatic channel selection results. Maximising the signal to noise ratio results in a selection with most channels close to the electrodes with highest peak magnitude. Note that some channels far away from the spike with possibly little correlation to the spike activity are selected which allows the algorithm to cancel some of the noise components.

4.2.6 Training scheme

Training and testing was performed individually for each subject. When evaluating the effect of inter-electrode distance, care was taken to separate the dataset in independent training and testing sets. The spike epochs were split in four folds of equal size. The rest of the non-spike data were also split in four folds each containing 30 minutes of data. Each training fold used three folds of spike epochs and three folds of non-spike data in order to estimate the covariance matrices $\mathbf{R_s}$ and $\mathbf{R_n}$. The obtained filter w defined in equation 4.4 was then evaluated on the remaining spike fold and remaining non-spike fold. This process was repeated four times.

4.2.7 Evaluation metrics

Since the real ground truth is unknown, we propose several indirect metrics to assess the quality of the spike detections.

Signal-to-Noise Ratio

The average spike signal power over 200 ms was computed as the average RMS amplitude over all channels in all segments that were flagged as spikes by Persyst (note that we used Persyst in the high specificity setting, such that the effect of the false positive detections are assumed to be negligible). The noise power was computed in a similar fashion over all non-spike segments. The ratio of both RMS values was then used as the SNR metric.

Correlation

To compare the Persyst detections and the detections of our max-SNR pipeline, a correlation score was computed for each subject between the average spike events detected by both algorithms. The correlation score allows to compare the morphology of the detected spike waveforms of both algorithms. A high correlation indicates that the number of false positive detections of the max-SNR filter is small (as these would have a distorting effect on the waveform). Indeed, since Persyst is used in the low sensitivity mode [143] (resulting in a high specificity, i.e., very few false positives), the average waveform of both approaches should be similar. The events were first aligned to the maximum of the absolute value over a 0.5 second epoch across all channels. The average 0.5 second event was then constructed. The Pearson-correlation coefficient was

computed per channel between the average of the spikes detected by Persyst and the spikes annotated by the max-SNR pipeline. The average correlation coefficient was computed as the weighted sum over all channels where the weight was given by the power of the channel in the Persyst average spike event.

Agreement score

Cohen's kappa coefficient of agreement was used as the agreement score. It was computed between the baseline performance of the max-SNR detection pipeline when trained with 32 channels referenced to `Cz` and between the different short inter-electrode distance scenarios. The inter-rater agreement between humans in annotating IEDs reaches a Cohen's kappa of 0.49 for detecting individual spikes [142]. We therefore consider a kappa larger than 0.5 to be in the region of practical equivalence (ROPE).

4.2.8 Statistical analysis

A Bayesian analysis was performed to evaluate the probability of obtaining a Cohen's kappa coefficient > 0.5 for the different short distance scenarios (when comparing with the baseline). This is the level of agreement obtained between human experts annotating interictal spikes [142]. The analysis was based on the generation of a large number of samples from distributions that fit to the observed data. A Bayesian analysis allows to estimate the probability of observing a parameter based on observed measurements and a prior assumption on the underlying distribution of the parameter. The observed Cohen's kappa was modeled as a $\mathcal{B}$-distribution as is appropriate for a variable bound to [0, 1] [147]. The $\mathcal{B}$-distribution was expressed in terms of mean and variance as was proposed in [148]. The mean and variance are set by a prior distribution with fixed parameters and are used to generate different distributions of Cohen's kappa coefficient. As proposed in [149], the means were drawn from a $\mathcal{B}$-distribution whose mean and variance was set to the mean and variance of Cohen's kappa coefficients in the full data set (i.e. across different subjects, folds, inter-electrode distances, and number of EEG units).. The variance was drawn from a uniform distribution in the interval between 0 and two times the variance of the whole data. The model is represented schematically in figure 4.2.4. PyMC 4.1.4 [150] was used to generate the model.

The complete code for the simulations is provided online : `https://github.com/danjjl/miniEEG`

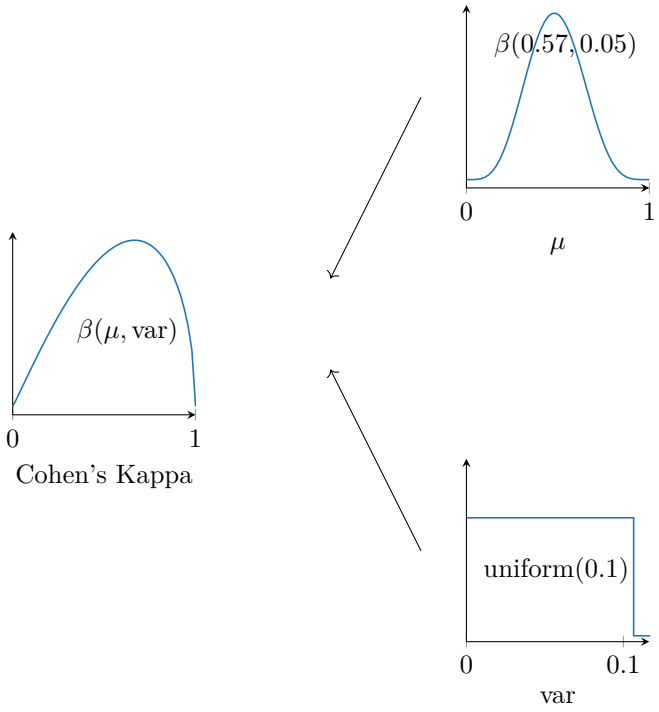


Figure 4.2.4: Schematic representation of the Bayesian model. Cohen's kappa coefficient is modeled as a β distribution with parameters obtained from a β distribution for the mean (μ) and a uniform distribution for the variance (var).

4.3 Results

4.3.1 Signal power

The power of the spike events annotated by Persyst is reported as a function of inter-electrode distance. The power is reported as an average across channels for the channels selected by the channel selection algorithm when 1-10 channels are selected. Figure 4.3.1.a shows that the median spike RMS at an inter-electrode distance of 2 cm is 4 μV and increases to 12 μV at 8 cm. The RMS is shown as a function of distance in 4.3.1.b. As opposed to the spike SNR, It shows a near equal SNR independent of distance with a median SNR of 1.4 at 2 cm and 1.5 at 8 cm. This implies that both the power of the spikes as well as the background noise are reduced by an equal amount when reducing the

inter-electrode distances.

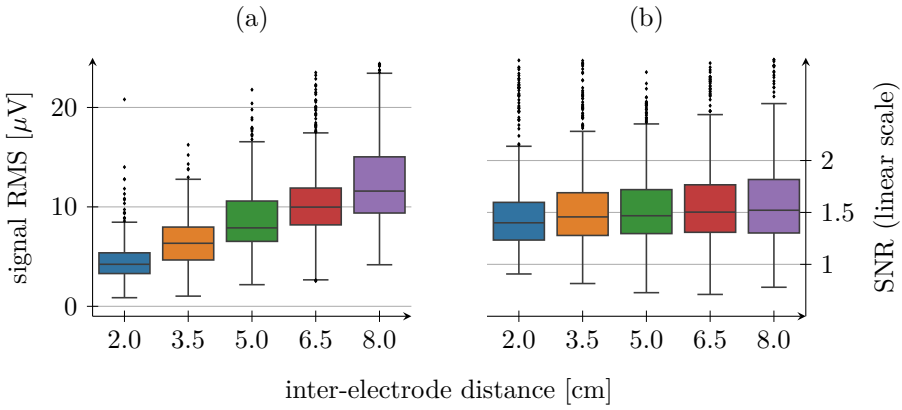


Figure 4.3.1: Boxplots of (a) spike RMS amplitude as a function of distance and (b) spike RMS to noise RMS ratio as a function of distance over 15 subjects and 1-10 EEG channels.

4.3.2 Correlation between Persyst and the max-SNR pipeline

In order to validate our computationally efficient max-SNR based spike detection pipeline, we compared the spikes detected by Persyst to the spikes detected by the max-SNR based detection pipeline when electrodes were referenced to `Cz` and 32 channels were selected using the automatic channel selection algorithm. Both algorithms detect events such that the average spike waveform of both algorithms have a very high correlation. The distribution of average spike correlations is given in figure 4.3.2.b. It shows a median correlation of 0.94. An example of the average spike of both algorithms for subject 15 is given in 4.3.2.a. The number of events detected by the max-SNR algorithm is four times higher than the number of events Persyst detected at the low sensitivity setting. In total, across all subjects, the max-SNR algorithm detected 16937 events while Persyst detected 4088.

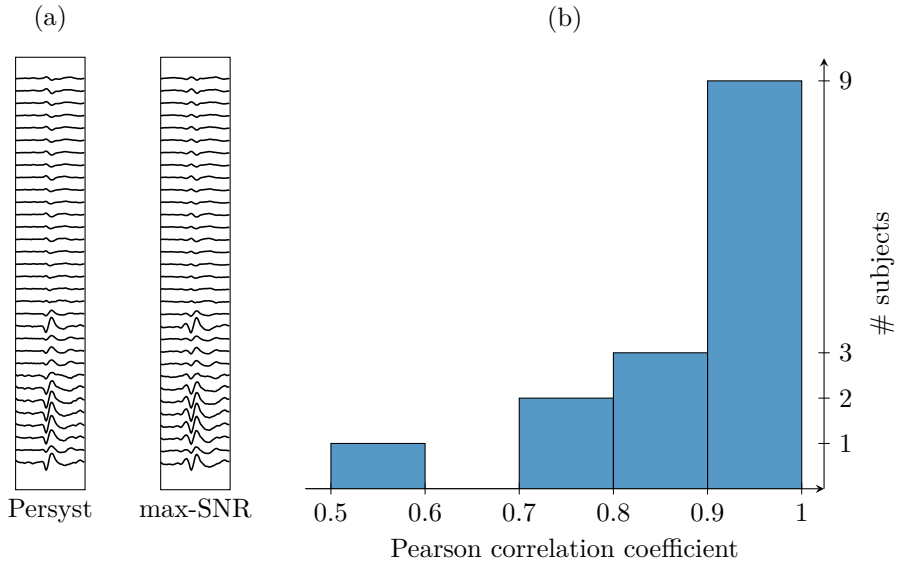


Figure 4.3.2: Comparison of the average spike event from patient 15 detected by Persyst and by the max-SNR pipeline. (a) Example average spike event : (left) Persyst, (right) max-SNR pipeline. One second of data is shown with 32 electrodes referenced to `Cz`. (b) Histogram across all patients of the Pearson correlation between the average spike event of both algorithms.

4.3.3 Cohen's kappa coefficient as a function of number of EEG units and distance

Cohen's kappa coefficient of agreement is computed between the max-SNR spike detections on 32 channels referenced to `Cz` and max-SNR spike detections when the number of channels was restricted between one and ten and the inter-electrode distance was fixed to 2, 3.5, 5, 6.5, 8 cm. A boxplot of these different comparisons is shown in figure 4.3.3. The median of Cohen's kappa coefficient increases both with the number of units and with the inter-electrode distance. For an inter-electrode distance of 2 cm, the median coefficient is always below 0.5 for three or less nodes. From 3.5 cm, it increases above 0.5 for two or more EEG units.

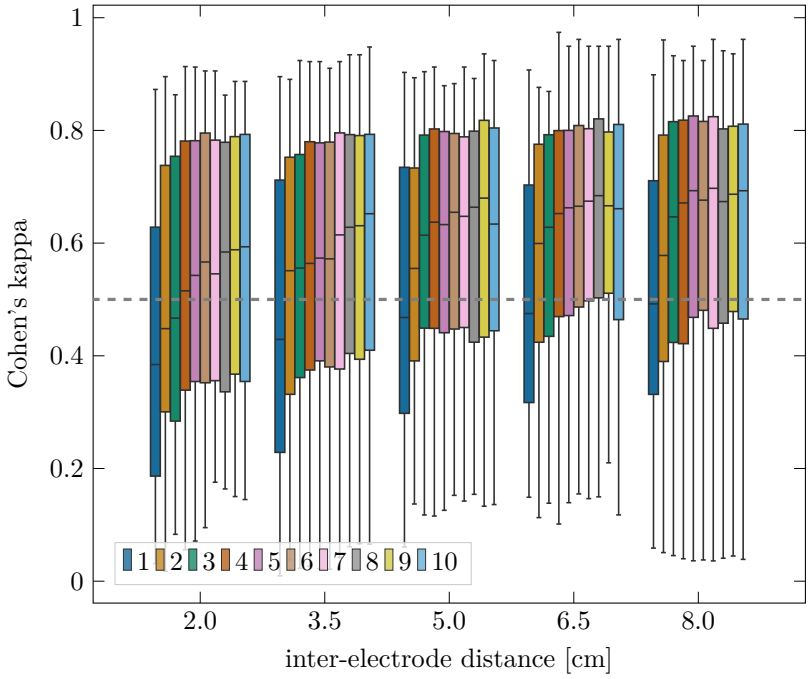


Figure 4.3.3: Boxplot of Cohen's kappa coefficient of agreement of detected spike events between the different test scenarios (varying distance and number of channels) using a max-SNR pipeline and a reference max-SNR pipeline on 32 channels referenced to `Cz`. The different colors represent different number of EEG units. The dashed grey line shows the agreement between human raters.

The Bayesian framework is used to test the confidence of these findings. All simulated MCMC reported convergence with the Gelman-Rubin statistic equal to one for all variables. The probability of the mean Cohen's kappa coefficient being greater than 0.5 is reported in figure 4.3.4. It shows a probability $< 95\%$ when using a single EEG unit. It shows a probability $> 95\%$ when using seven or more EEG units at distances of 2 and 3.5 cm. It shows a probability $> 95\%$ when using two or more EEG units at distances of 5 cm or more.

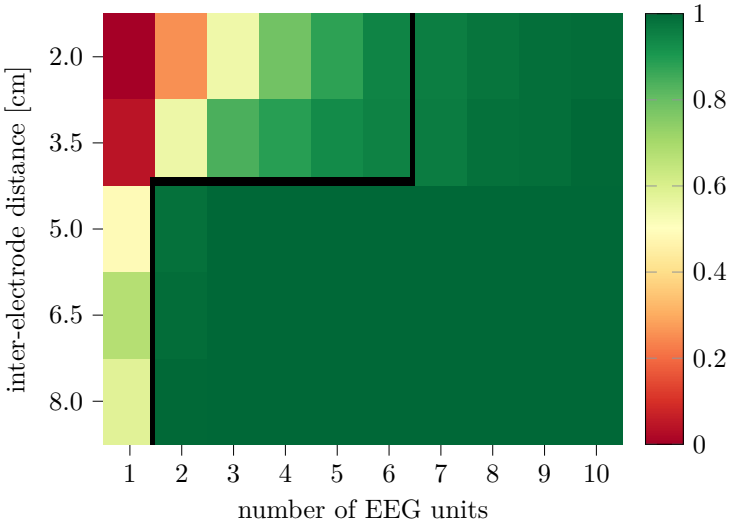


Figure 4.3.4: Probability in a Bayesian model of observing a mean Cohen's kappa coefficient of agreement > 0.5 between spike events detected in different test scenarios (varying distance and number of EEG units) using a max-SNR pipeline and a reference max-SNR pipeline on 32 channels referenced to `Cz`. The black line shows the 95% probability boundary.

4.4 Discussion

In this chapter, we investigated the effect of inter-electrode distance and influence of the number EEG channels on the ability to record and detect spikes using a set of galvanically isolated miniature EEG sensor units. We analyzed data recorded for clinical purposes with a 257 electrode HD-EEG setup. We annotated the data for spikes using the Persyst 14.D tool, which has been widely utilized in clinical practice and research for seizure and spike detection on EEG. We then measured the RMS amplitude of the detected spikes after re-referencing to short inter-electrode distances. We found that the median RMS of spike events measured with an inter-electrode distance of 2 cm was 4 μV (see figure 4.3.2.A). This is about three times lower signal power than found at greater electrode distances and results in signals that are close to the noise floor of standard EEG equipment. In standard clinical care, the American Clinical Neurophysiology Society states that EEG recording devices should add less than 1 μV peak-to-peak noise at any frequency in the band [0.5 - 100] Hz [151]. This is in line with amplifier technology commonly found in EEG equipment,

such as the Texas Instrument ADS1299 which claims an input referred-noise of 1 μV peak-to-peak [152]. This indicates that EEG devices with short inter-electrode distances should aim for lower input referred noise. This is a similar observation as in [153]. Our findings suggest that devices aiming for 2 cm inter-electrode distance should be designed for an input referred noise < 300 nV (three times lower than current standard as suggested by spike amplitude). We also compared the spike RMS to the non-spike RMS (SNR) as a function of inter-electrode distance. This SNR results both from physiological processes as well as extra-physiological such as electronic noise. We found a ratio of 150%. This is in line with the definition of spikes, which state that spikes should stand out from the background EEG activity [19]. The SNR was only minimally impacted by the inter-electrode distance with the SNR at 140% for an inter-electrode distance of 2 cm. This indicates that spike activity is still distinguishable at 2 cm with scalp EEG. This also means that the amplitude of the spikes and the noise (i.e. background non-spike EEG) is reduced with a similar factor. A similar observation was also made in [153] for the case of event-related potentials.

We compared the spike annotations from the Persyst software on the original 257 electrodes to the annotations using our channel selection and detection pipeline constrained to 32 channels referenced to `Cz`. We found that Persyst detected about four times less spikes than our proposed algorithm. However, we observed a median Pearson correlation of 0.94 between the average spike event detected by both algorithms, with a single subject with a correlation of less than 0.77. For this subject, Persyst does not detect a clear spike pattern. The maximum amplitude of the average spike is 9 μV with no clear spike morphology. The high median correlation indicates a very good agreement on the morphology of spikes between both algorithms and is a point of validation to demonstrate that our algorithm does detect valid spike events. This indicates that the additional spikes detected by the max-SNR filter are actual spikes instead of false positives, implying that Persyst misses a large fraction of the spike events. The latter is expected, since we set Persyst to low sensitivity mode because this is the mode with the highest specificity. In this setting, Reus et al. observed a specificity of 99% when compared to human experts [143]. This setting allows us to use Persyst as reliable ground truth data to accurately estimate the $\mathbf{R_s}$ covariance matrix with minimal leakage of non-spike events in the estimation of $\mathbf{R_s}$.

We compared the spike annotations between a baseline version of our detection pipeline referenced to `Cz` in a 32 channel setting with the annotations from our pipeline constrained to short inter-electrode distances and less than ten channels. We found an influence of both the distance and the number of EEG units on Cohen's kappa coefficient of agreement. Increasing distance increases agreement. Increasing the number of EEG units also increases agreement. We

found that the influence of the number of EEG units is stronger at small inter-electrode distances. The increase in performance is strongest when taking two or more units compared to a single EEG unit. We used the Bayesian estimation framework [149] to estimate the probability of obtaining high agreement. This Bayesian testing confirmed a probability $> 95\%$ of obtaining an agreement > 0.5 for inter-electrode distances greater than 5 cm when using two or more EEG units (which corresponds to an agreement that is at least as good as an inter-rater agreement of human annotators, for which Cohen's kappa is 0.49 [142]). These findings indicate that a network of galvanically isolated short distance EEG units is an appropriate setup to monitor spikes and is comparable to a 32 channel setup referenced to `Cz`. It also shows that carefully placed units and increasing the number of units can counteract the decrease in performance obtained by shortening the inter-electrode distance.

Current efforts to miniaturize EEG systems are not limited to the field of epilepsy. In the hearing aid community, several research groups have investigated miniaturization of EEG systems to integrate them with hearing aids. In a recent study, Mundanad Narayanan et al. showed that for inter-electrode distances equal or greater to 3 cm, the decoding of neural responses to speech was not significantly worse to that achieved with long distance channels referenced to the `Cz` electrode [47] when using an appropriate number of EEG units. While a direct comparison to a spike detection task is impossible, given the differences in the nature of the neural response localization and amplitude, both our study and the studies in the hearing aid community show networks of galvanically isolated short distance EEG units are able to capture an EEG signal of interest.

The design objective of short distance EEG units is to allow for measurement of EEG in real-life conditions. This study was conducted on data acquired in a hospital setting. Recordings obtained in uncontrolled conditions during everyday life are expected to contain more artifacts than measured in the controlled setting of the epilepsy monitoring unit. Therefore, reproducibility of the findings we report on a minimal number of channels and inter-electrode distance should now be tested in everyday life activity conditions. When designing an ambulatory setup, robustness should be considered. This could be done by increasing the number of channels and/or inter-electrode distance above our minimal requirements findings. Improvements to the spike detection algorithm could also be made for example by building a filter that discriminates between epileptic spikes and non-epileptic peak interferers as was suggested in [104].

4.5 Conclusion

We found that EEG recording equipment should be specifically designed to measure the small signal power at short inter-electrode distance by providing an input referred noise < 300 nV. We also found an inter-electrode distance of 5 cm between electrodes in a setup with a minimum of two EEG units is required to obtain near equivalent performance in interictal spike detection to standard EEG. These findings provide design guidelines for EEG equipment miniaturization in the context of detection of interictal epileptic discharges.

CHAPTER 5

Conclusion

This work contributes to the development of EEG monitoring of people with epilepsy in the home environment and lowering the technical difficulties that limit its wide adoption. It brings three main contributions which are detailed in the main chapters of this dissertation. This chapter will summarise these contributions as well as propose future directions for research in the field.

5.1 Main contributions

Epileptiform event detection | The first contribution of this thesis lies in the development and validation of a computationally efficient algorithm for automatic detection of epileptiform events in wearable EEG.

It is designed for systems using wearable EEG sensors for long-term monitoring, e.g. over 24 hours or multiple days of EEG recording. Such recordings generate large amounts of data that need to be analysed to extract relevant physiological

information. Annotation of epileptic seizures on EEG is a time-consuming task that is performed by a highly trained epileptologist. It would be greatly advantageous that the task would be automated for use with wearable EEG. In addition, certain applications of wearable EEG, such as seizure alarms, require algorithms in real-time on the wearable EEG device. In contrast to algorithms developed to run in a hospital without strict computational constraints (as they can rely on powerful servers connected to the hospital network) [70], real-time algorithms for wearable systems must meet strict storage, computing memory, and computing power constraints. These real-time algorithms can be used to provide an alarm to a caregiver when a seizure is detected or could be used as a trigger to provide abortive seizure medication. In chapter 2, we design an algorithm for real-time absence seizure detection specifically for use on a microcontroller or field-programmable gate arrays (FPGA) such as found in wearable EEG. The algorithm is based on a spatio-temporal filter that maximally separates absence seizures from peak interferers. The design of this filter builds on previously reported max-SNR filters, which are themselves analogous to matched filters. The filter we propose adds several key innovations. While max-SNR maximally separate between a target (seizure samples) and noise (non-seizure samples), the max-seizure to peak interference ratio (SPIR) filter we propose is able to achieve better classification outcome by separating between seizures and peak-interferers (i.e. non-seizure samples with high power) because low power 'background EEG' are already easily distinguishable from seizures. We also propose a regularisation scheme for the max-SPIR filter based on existing work in template matching. This regularisation greatly improves the performance of the filter when presented with data covariance matrices that are ill-conditioned due to a lack of training data. In chapter 2, the algorithm was benchmarked against two state-of-the-art absence seizure detection algorithms (without computation constraints), and was found to perform almost on par with the best of these when used in a configuration with ten or more channels, while being much more efficient in terms of hardware memory and computational requirements. The algorithm demonstrates the relevance of a GEVD based filtering pipeline for pattern identification in epilepsy. We used this same approach to identify IEDs in chapter 4. In that chapter we showed that the GEVD-based filtering detected similar IED events to those detected by a commercial software.

In addition to using GEVD-based filtering for the detection of epileptiform events in low power devices, it can also be used to achieve state of the art seizure detection in computationally intensive pipelines. In "Epileptic Seizure Detection in EEG via Fusion of Multi-View Attention-Gated U-Net Deep Neural Networks" [55], together with colleagues, we proposed a seizure detection algorithm based on multiple deep neural networks, each trained on a different pre-processing pipeline. One of these pre-processing pipelines was based on GEVD-based

filtering. It was shown that including this GEVD-based filtering improved the overall classification.

Chapter 2 showed that 10 well-chosen channels were sufficient to reach 0.5 false positives per day for an absence seizure detection sensitivity of 95% in all but one patient. This highlights both the feasibility of using a reduced number of channels compared to the clinical gold-standard in the context of seizure monitoring with wearable EEG as well as the need for a good channel selection procedure.

Electroencephalography channel selection | The second contribution proposes a new channel selection algorithm as well as a benchmark of state-of-the-art sensor selection methods in GEVD problems. Channel selection is essential when designing a wearable EEG system as these should only use a minimum number of electrodes to optimise portability. The methods we propose are data-driven, i.e. the channel selection is based on optimising the GEVD-based epileptiform event detection task. It can be generalised as a variable selection task and is an important problem in many mathematical engineering fields. For example, the variable selection problem is found in telecommunication, where antenna placement is critical to the good functioning of a communication network, biomedical sensor arrays, or wireless acoustic sensor networks, where a microphone subset needs to be selected.

In chapter 3, we propose a new channel selection algorithm. It extends the GEVD channel selection for purely spatial filtering in [108] to spatio-*temporal* filtering borrowing techniques from [107]. This necessitates the use of a *group-sparse* regulariser. When a channel is eliminated, not just one variable, but all corresponding filter lags should be put to zero. The algorithm provides the possibility to take multiple filters (i.e. multiple generalised eigenvectors) into account (multiple-input multiple-output (MIMO)), whereas previous work focused on the dominant generalised eigenvector (multiple-input single-output (MISO)). This requires consistent removal or zeroing of the filter coefficients corresponding to an eliminated channel across *all* filters. This approach can be employed in various other applications, where the notion of a shared selection exists. In the chapter we also identify a key failure of the otherwise performant and computationally efficient background elimination method. This method fails when the selection is performed on ill-conditioned data matrices. We proposed and demonstrated the efficacy of a regularisation scheme for the backward elimination method. Finally, the chapter brings about evidence which had been largely missing in prior literature, namely an in-depth and statistical comparison of state-of-the-art channel selection methods in GEVD problems.

Limits of EEG miniaturisation | This thesis provides an analysis of the limits of miniaturisation of wearable EEG systems. It analyses the effect of reducing the number of channels (or EEG units in a network of galvanically isolated EEG sensors) as well as the effect of inter-electrode distance in a setup with fixed inter-electrode distance for all channels. This analysis is conducted by emulating short distance EEG on high-density EEG recordings with 257 electrodes. Chapter 4 made several key observations for the design of wearable EEG. First it quantified the power of IED events in function of inter-electrode distance. This allowed the definition of minimum required specifications for the acquisition of IED at short inter-electrode distance. The chapter also shows the limits of accurate IED detection when reducing both the number of EEG channels and the inter-electrode distance. This allows for better design specifications when building new wearable EEG systems.

In summary | This thesis contributes to the development of wearable EEG. It provides three major contributions to the field :

- An automated epileptiform event detection algorithm based on a computationally efficient GEVD-based filter. (chapters 2,4)
- A channel selection algorithm for GEVD problems as a well as a comparison with state-of-the-art channel selection methods in GEVD problems. (chapter 3)
- An analysis of the limits of miniaturisation of EEG systems for detecting IED. (chapter 4)

5.2 Open challenges & Future perspectives

The work in this PhD thesis contributes to the development of wearable EEG. There remain some open challenges related to this body of work. In this section, we will give an overview of these challenges and we will sketch some ideas that could serve as a starting point for future research.

The GEVD-based filtering presented in this dissertation was successfully applied to IED detection collected from in-hospital EEG recordings. We also successfully applied the method to absence seizure detection in EEG recorded in people while engaged in their everyday life without any movement constraints. The filters proposed in this thesis make the assumption that the signal is stationary. However, the stationarity assumption is easily violated in practice. Several causes can result in non-stationarity of the signal statistics. Among them,

changes in electrode impedance lead to degradation of the signal quality and changes in the signal statistics. Changes in brain states could also influence the signal statistics, but should only have a minimal impact on the filter performance if training data was presented with these different states. Everyday life activities can also lead to changes in the statistics of the signal. Repetitive artefacts such as those found while walking or vibration associated with motorised transportation can lead to non-stationarity of the EEG. It is not expected that small changes in the signal statistics would significantly degrade the performance of the detection pipeline. However, we can assume that an adaptive filter that would update in function of the statistics of the signal would improve the detection performance. The following strategy proposed in [154] could be used to build an adaptive filter :

1. Estimate the seizure ($\mathbf{R}_{ss}$) and interference ($\mathbf{R}_{ii}$) covariance matrix based on some annotated training data.
2. Design an initial max-SPIR filter that discriminates between the seizures and interferers.
3. During the inference phase, every filtered epoch that crosses a threshold for inclusion as an interferer is used to update the interferer covariance ($\mathbf{R}_{ii}$) matrix. This update can be done with an exponential forgetting factor which slowly favours new observations.
4. Also during the inference phase, every detected seizure can be used to update the seizure covariance matrix ($\mathbf{R}_{ss}$) estimate.
5. These updated covariance matrices can then be used to update the max-SPIR filter on a regular basis.

The stability of this updating mechanism would have to be evaluated to make sure the updates based on unlabelled data do not lead to serious degradation of the performance of the filter. A computationally efficient mechanism to invert the interferer covariance matrix and compute the filter would also have to be designed.

The epileptiform event detection pipeline developed in this thesis was evaluated both on IED detection collected from in-hospital EEG recordings of patients with focal epilepsy and on absence seizures recorded in people while engaged in their everyday life without any movement constraints. Both these epileptiform activities present stereotypical spatio-temporal patterns that can be well discriminated from noise by the GEVD-based filtering. Other seizure types might require further algorithmic development. These include generalised tonic-clonic seizures, which are often accompanied by strong movement artefacts,

as well as focal seizures, which typically evolve over the course of the seizure [155]. For these seizure types, the GEVD-based filtering would likely not have optimal performance. Algorithms for focal seizure detection would likely need to integrate some of the dynamic features of the seizures. This could possibly be done by using a filterbank of GEVD-based filters informed by a hidden Markov model [156].

In this thesis we designed a centralised version of the GEVD-based filter. This means a central processing unit needs access to the multi-channel EEG signals to compute the filter. In a practical implementation, this would require communication between the miniaturised EEG nodes. The communication with a central unit could be wired, but this would severely hinder the discreetness and user-comfort of the device. It could be wireless, and this would severely impact battery life [44]. An alternative, energy efficient, approach would rely on distributed processing. This implies that one cannot measure the full covariance structure across all EEG channels. To this end, algorithms that iteratively perform in-network fusion where the EEG units only exchange fused data with a few (1-hop) neighbours would have to be designed. Over time, all nodes gradually adapt their fusion filters driven by the (fused) data that pass through them, thereby keeping track of local correlation matrices. The distributed realisation of the algorithms proposed in this thesis could rely on existing literature, such as the work presented by Bertrand et al. [48], [157].

The filter we propose produces a single channel output corresponding to the eigenvector associated to largest eigenvalue. This single channel output is the signal with highest SNR. However, the GEVD-based filtering could produce multiple filter outputs. These multiple outputs could then be used to extract a feature vector for a classifier. This approach is used in the brain computer interface (BCI) community when using common spatial patterns (CSP) filters [109]. In the context of epileptiform activity detection, multiple filter outputs could provide substantial performance improvements when multiple distinct epileptiform patterns should be detected. This would for example be the case for people with two distinct clusters of IED and it might be the case in less stereotypical seizures such as those found in some patients with focal seizures.

This thesis focused on signal processing problems related to wireless EEG sensor networks for monitoring people with epilepsy. Other challenges also need to be resolved for wider adoption of the technology. In this thesis, all data collected in chapter 2 and in chapter 4 were obtained with gold-cup electrodes and gel or wet sponges. These wet electrode systems dry over time and cannot be used for several days (up to 24 hours) without manual intervention to refill the gel and check electrode impedance. The gold-cup electrodes used in this study also required special training to manipulate. This requires an EEG technician to place the electrodes and periodically check the electrode impedance. For wider

adoption of the technology, the electrode should be easy enough to manipulate for a user to place them himself. Studies such as Plug 'n Patch [158] are testing electrode adhesives that should provide solutions to this problem, while microneedles and dry electrode technology presented in the introduction could enable long term monitoring.

The system of wireless EEG sensor networks would also have to be further evaluated in the field. In this PhD thesis only a limited subset of recordings were performed in people that were allowed to continue with their everyday life activities. It is expected that EEG recorded in a free environment would be more heavily affected by artefacts compared to EEG recorded in the clinic. This will require adequate signal processing to distinguish signal of interest from noise. The algorithm here should provide a partial solution. However, it is possible that additional pre-processing steps would be required to add to the robustness of the system for in the event an electrode disconnects during the recording. Other ways of dealing with the robustness of the system could be achieved by using different sensing modalities to monitor epilepsy. Studies have shown that in selected patients, seizure classification could be improved by adding accelerometery, electromyography (EMG) or electrocardiography (ECG) [159], [160].

In sum, this PhD thesis introduces several new signal processing methods for wireless EEG sensor networks for monitoring of people with epilepsy. It contributes to the technological advancement required for the wider adoption of this technology.

Bibliography

[1] R. S. Fisher, C. Acevedo, A. Arzimanoglou, *et al.*, "ILAE official report: A practical clinical definition of epilepsy", *Epilepsia*, vol. 55, pp. 475–482, 4 2014.

[2] V. L. Feigin, E. Nichols, T. Alam, *et al.*, "Global, regional, and national burden of neurological disorders, 1990-2016: A systematic analysis for the global burden of disease study 2016", *The Lancet. Neurology*, vol. 18, pp. 459–480, 5 2019.

[3] K. M. Fiest, K. M. Sauro, S. Wiebe, S. B. Patten, C. S. Kwon, J. Dykeman, T. Pringsheim, D. L. Lorenzetti, and N. Jetté, "Prevalence and incidence of epilepsy: A systematic review and meta-analysis of international studies", *Neurology*, vol. 88, pp. 296–303, 3 2017.

[4] E. Beghi, "The epidemiology of epilepsy", *Neuroepidemiology*, vol. 54, pp. 185–191, 2 2020.

[5] I. E. Scheffer, S. Berkovic, G. Capovilla, *et al.*, "ILAE classification of the epilepsies: Position paper of the ILAE commission for classification and terminology", *Epilepsia*, vol. 58, pp. 512–521, 4 2017.

[6] S. T. Sarmast, A. M. Abdullahi, and N. Jahan, "Current classification of seizures and epilepsies: Scope, limitations and recommendations for future action", *Cureus*, vol. 12, 9 2020.

[7] C. P. Panayiotopoulos, "General aspects of epilepsies", in *A Clinical Guide to Epileptic Syndromes and their Treatment*, 2nd edition, Springer, 2010, pp. 1–19.

[8] T. Proix, W. Truccolo, M. G. Leguia, T. K. Tcheng, D. King-Stephens, V. R. Rao, and M. O. Baud, "Forecasting seizure risk in adults with focal epilepsy: A development and validation study", *The Lancet Neurology*, vol. 20, pp. 127–135, 2 2021.

[9] R. E. Stirling, M. J. Cook, D. B. Grayden, and P. J. Karoly, "Seizure forecasting and cyclic control of seizures", *Epilepsia*, vol. 62, S2–S14, S1 2021.

[10] C. Meisel and T. Loddenkemper, "Seizure prediction and intervention", *Neuropharmacology*, vol. 172, 2020.

[11] J. J. Riviello, S. Ashwal, D. Hirtz, T. Glauser, K. Ballaban-Gil, K. Kelley, L. D. Morton, S. Phillips, E. Sloan, and S. Shinnar, "Practice parameter: Diagnostic assessment of the child with status epilepticus (an evidence-based review)", *Neurology*, vol. 67, pp. 1542–1550, 2006.

[12] D. M. Ritter and K. Holland, "Genetic testing in epilepsy", *Seminars in neurology*, vol. 40, pp. 730–738, 6 2020.

[13] S. R. Benbadis, S. Beniczky, E. Bertram, S. MacIver, and S. L. Moshé, "The role of eeg in patients with suspected epilepsy", *Epileptic disorders*, vol. 22, pp. 143–155, 2 2020.

[14] H. Hinrichs, M. Scholz, A. K. Baum, J. W. Kam, R. T. Knight, and H. J. Heinze, "Comparison between a wireless dry electrode eeg system with a conventional wired wet electrode eeg system for clinical applications", *Scientific Reports*, vol. 10, pp. 1–14, 1 2020.

[15] S. R. Sinha, L. Sullivan, D. Sabau, D. San-Juan, K. E. Dombrowski, J. J. Halford, A. J. Hani, F. W. Drislane, and M. M. Stecker, "American clinical neurophysiology society guideline 1: Minimum technical requirements for performing clinical electroencephalography", *Journal of Clinical Neurophysiology*, vol. 33, pp. 303–307, 4 2016.

[16] G. H. Klem, H. O. Lüders, H. H. Jasper, and C. Elger, "The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology", *Electroencephalogr Clin Neurophysiol Suppl*, vol. 52, pp. 3–6, 1999.

[17] V. Jurcak, D. Tsuzuki, and I. Dan, "10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems", *NeuroImage*, vol. 34, pp. 1600–1611, 4 2007.

[18] D. L. Schomer and F. H. Lopes da Silva, *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Oxford University Press, 2017.

[19] M. A. Kural, L. Duez, V. S. Hansen, *et al.*, "Criteria for defining interictal epileptiform discharges in EEG", *Neurology*, vol. 94, e2139–e2147, 20 2020.

[20] D. B. Burkholder, J. W. Britton, V. Rajasekaran, *et al.*, "Routine vs extended outpatient eeg for the detection of interictal epileptiform discharges", *Neurology*, vol. 86, p. 1524, 16 2016.

[21] G. P. Kalamangalam, "Extracranial interictal and ictal eeg in seeg planning", *Neurosurgery clinics of North America*, vol. 31, pp. 345–371, 3 2020.

[22] S. J. M. Smith, "EEG in the diagnosis, classification, and management of patients with epilepsy", *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 76, pp. ii2–ii7, suppl 2 2005.

[23] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment", Ph.D. dissertation, Harvard University, 2009.

[24] M. Koutroumanidis, A. Arzimanoglou, R. Caraballo, *et al.*, "The role of EEG in the diagnosis and classification of the epilepsy syndromes: A tool for clinical practice by the ILAE neurophysiology task force (part 1)", *Epileptic Disorders*, vol. 19, pp. 233–298, 3 2017.

[25] C. Miskin, K. S. Carvalho, I. Valencia, A. Legido, and D. S. Khurana, "EEG duration: The long and the short of it", *Journal of Child Neurology*, vol. 30, pp. 1767–1769, 13 2015.

[26] C. Mann, L. M. Willems, A. C. Leyer, T. M. Freiman, J. Konczalla, M. Kieslich, F. Rosenow, A. Strzelczyk, and S. Schubert-Bast, "Benefits, safety and outcomes of long-term video EEG monitoring in pediatric patients", *European Journal of Paediatric Neurology*, vol. 32, pp. 29–35, 2021.

[27] W. O. Tatum, J. J. Halford, P. Olejniczak, *et al.*, "Minimum technical requirements for performing ambulatory EEG", *Journal of clinical neurophysiology*, vol. 39, 6 Jul. 2022.

[28] D. Velis, P. Plouin, J. Gotman, and F. L. D. Silva, "Recommendations regarding the requirements and applications for long-term recordings in epilepsy", *Epilepsia*, vol. 48, pp. 379–384, 2 2007.

[29] U. Seneviratne, A. Mohamed, M. Cook, and W. D'Souza, "The utility of ambulatory electroencephalography in routine clinical practice: A critical review", *Epilepsy Research*, vol. 105, pp. 1–12, 1-2 2013.

[30] Micromed. "BRAIN QUICK Ambulatory - Micromed Group". (2022), [Online]. Available: https://micromedgroup.com/brainquick/brainquick-ambulatory/ (visited on 09/28/2022).

[31] F. Brunnhuber, J. Slater, S. Goyal, D. Amin, G. Thorvardsson, D. R. Freestone, and M. P. Richardson, "Past, present and future of home video-electroencephalographic telemetry: A review of the development of in-home video-electroencephalographic recordings", *Epilepsia*, vol. 61, S3–S10, S1 2020.

[32] T. U. Syed, W. C. LaFrance, T. Loddenkemper, *et al.*, "Outcome of ambulatory video-EEG monitoring in a 10,000 patient nationwide cohort", *Seizure*, vol. 66, pp. 104–111, 2019.

[33] N. S. Williams, G. M. McArthur, B. de Wit, G. Ibrahim, and N. A. Badcock, "A validation of emotiv EPOC flex saline for EEG and ERP research", *PeerJ*, vol. 8, e9713, 2020.

[34] O. E. Krigolson, M. R. Hammerstrom, W. Abimbola, R. Trska, B. W. Wright, K. G. Hecker, and G. Binsted, "Using muse: Rapid mobile assessment of brain performance", *Frontiers in Neuroscience*, vol. 15, p. 56, 2021.

[35] L. Swinnen, C. Chatzichristos, K. Jansen, *et al.*, "Accurate detection of typical absence seizures in adults and children using a two-channel electroencephalographic wearable behind the ears", *Epilepsia*, vol. 62, pp. 2741–2752, 11 2021.

[36] G. Stavrinidis, K. Michelakis, V. Kontomitrou, G. Giannakakis, M. Sevrisarianos, G. Sevrisarianos, N. Chaniotakis, Y. Alifragis, and G. Konstantinidis, "Su-8 microneedles based dry electrodes for electroencephalogram", *Microelectronic Engineering*, vol. 159, pp. 114–120, 2016.

[37] R. Wang, X. Jiang, W. Wang, and Z. Li, "A microneedle electrode array on flexible substrate for long-term EEG monitoring", *Sensors and Actuators B: Chemical*, vol. 244, pp. 750–758, 2017.

[38] M. Vafaei, A. Parhizgar, E. Abiri, and M. R. Salehi, "A low power and ultra-high input impedance analog front end based on fully differential difference inverter-based amplifier for biomedical applications", *AEU - International Journal of Electronics and Communications*, vol. 142, p. 154 005, 2021.

[39] S. Debener, R. Emkes, M. De Vos, and M. Bleichner, "Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear", *Sci Rep*, vol. 5, p. 16 743, 2015.

[40] S. L. Kappel, M. L. Rank, H. O. Toft, M. Andersen, and P. Kidmose, "Dry-contact electrode ear-eeg", *IEEE Transactions on Biomedical Engineering*, vol. 66, pp. 150–158, 1 2019.

[41] D. Sopic, A. Aminifar, and D. Atienza, "E-glass: A wearable system for real-time detection of epileptic seizures", *Proceedings - IEEE International Symposium on Circuits and Systems*, vol. May, 2018.

[42] Z. Gao, X. Cui, W. Wan, Z. Qin, and Z. Gu, "Signal quality investigation of a new wearable frontal lobe eeg device", *Sensors*, vol. 22, p. 1898, 5 2022.

[43] P. F. Viana, T. P. Attia, M. Nasseri, *et al.*, "Seizure forecasting using minimally invasive, ultra-long-term subcutaneous electroencephalography: Individualized intrapatient models", *Epilepsia*, 2022.

[44] A. Bertrand, "Distributed signal processing for wireless eeg sensor networks", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, pp. 923–935, 6 2015.

[45] M. Baijot, A. M. Narayanan, M. B. A. Rosa, J. Dan, A. Bertrand, and M. Kraft, "A miniature EEG node for synchronized wireless EEG sensor networks", in *Proceedings of 8th International Electronic Conference on Sensors and Applications*, MDPI, 2021.

[46] Y. M. Chi and G. Cauwenberghs, "Wireless non-contact EEG/ECG electrodes for body sensor networks", *2010 International Conference on Body Sensor Networks, BSN 2010*, pp. 297–301, 2010.

[47] A. M. Narayanan, R. Zink, and A. Bertrand, "EEG miniaturization limits for stimulus decoding with EEG sensor networks", *Journal of Neural Engineering*, vol. 18, 5 2021.

[48] C. A. Musluoglu and A. Bertrand, "A unified algorithmic framework for distributed adaptive signal and feature fusion problems – part i: Algorithm derivation", 2022.

[49] X. Jiang, G. B. Bian, and Z. Tian, "Removal of artifacts from eeg signals: A review", *Sensors*, vol. 19, 5 2019.

[50] Satyender, S. K. Dhull, and K. K. Singh, "A review on automatic epilepsy detection from EEGs signals", *Lecture Notes in Electrical Engineering*, vol. 668, pp. 1441–1454, 2021.

[51] M. K. Siddiqui, R. Morales-Menendez, X. Huang, and N. Hussain, "A review of epileptic seizure detection using machine learning classifiers", *Brain Informatics*, vol. 7, pp. 1–18, 1 2020.

[52] L. Wang, W. Xue, Y. Li, M. Luo, J. Huang, W. Cui, and C. Huang, "Automatic epileptic seizure detection in EEG signals using multi-domain feature extraction and nonlinear analysis", *Entropy*, vol. 19, p. 222, 6 2017.

[53] J. Jirka, M. Prauzek, O. Krejcar, and K. Kuca, "Automatic epilepsy detection using fractal dimensions segmentation and GP-SVM classification", *Neuropsychiatric Disease and Treatment*, vol. 14, pp. 2439–2449, 2018.

[54] A. Shoeibi, M. Khodatars, N. Ghassemi, *et al.*, "Epileptic seizures detection using deep learning techniques: A review", *International Journal of Environmental Research and Public Health*, vol. 18, p. 5780, 11 2021.

[55] C. Chatzichristos, J. Dan, A. M. Narayanan, N. Seeuws, K. Vandecasteele, M. D. Vos, A. Bertrand, and S. V. Huffel, "Epileptic seizure detection in eeg via fusion of multi-view attention-gated u-net deep neural networks", in *IEEE Signal Processing in Medicine and Biology Symposium (IEEE SPMB)*, J. Picone, Ed., IEEE, 2020.

[56] M. Aminoff, "Electroencephalography: General principles and clinical applications", in *Aminoff's Electrodiagnosis in Clinical Neurology*, M. J. Aminoff, Ed., Sixth, London: W.B. Saunders, 2012, ch. 3, pp. 37–84.

[57] E. Beghi, G. Giussani, E. Nichols, *et al.*, "Global, regional, and national burden of epilepsy, 1990-2016: A systematic analysis for the global burden of disease study 2016", *The Lancet Neurology*, vol. 18, no. 4, pp. 357–375, 2019.

[58] Y. W. Cho, G. K. Motamedi, and K. T. Kim, "The clinical utility of non-invasive video-electroencephalographic monitoring has been diversifying", *Neurological Sciences*, pp. 1–7, 2019.

[59] C. E. Elger and C. Hoppe, "Diagnostic challenges in epilepsy: Seizure under-reporting and seizure detection", *The Lancet Neurology*, vol. 17, no. 3, pp. 279–288, 2018.

[60] T. Kobulashvili, J. Höfler, J. Dobesberger, *et al.*, "Current practices in long-term video-EEG monitoring services: A survey among partners of the E-PILEPSY pilot network of reference for refractory epilepsy and epilepsy surgery", *Seizure*, vol. 38, pp. 38–45, 2016.

[61] C. P. Panayiotopoulos, "Epileptic seizures and their classification", in *A Clinical Guide to Epileptic Syndromes and their Treatment*. London: Springer London, 2010, pp. 21–63.

[62] H. Gastaut, "Part i: Definitions", in *Dictionary of Epilepsy*, Geneva: World Health Organization, 1973, ch. A, pp. 11–13.

[63] M. J. Keilson, W. A. Hauser, J. P. Magrill, and J. Tepperberg, "Ambulatory cassette EEG in absence epilepsy", *Pediatr. Neurol.*, vol. 3, no. 5, pp. 273–276, 1987.

[64] I. Zibrandtsen, P. Kidmose, C. Christensen, and T. Kjaer, "Ear-EEG detects ictal and interictal abnormalities in focal and generalised epilepsy – a comparison with scalp EEG monitoring", *Clinical Neurophysiology*, vol. 128, no. 12, pp. 2454–2461, 2017.

[65] F. Pinho, J. Cerqueira, J. Correia, N. Sousa, and N. Dias, "myBrain: A novel EEG embedded system for epilepsy monitoring", *Journal of Medical Engineering & Technology*, vol. 41, no. 7, pp. 564–585, 2017.

[66] D. Sopic, A. Aminifar, and D. Atienza, "e-Glass: A wearable system for real-time detection of epileptic seizures", in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, Florence, Italy, 2018, pp. 1–5.

[67] S. Boeckx, W. van Paesschen, B. Bonte, and J. Dan, "Live demonstration: SeizeIT - a wearable multimodal epileptic seizure detection device", in *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, Cleveland, Ohio, USA, 2018, pp. 1–1.

[68] A. Bach Justesen, M. T. Foged, M. Fabricius, C. Skaarup, N. Hamrouni, T. Martens, O. B. Paulson, L. H. Pinborg, and S. Beniczky, "Diagnostic yield of high-density versus low-density EEG: The effect of spatial sampling, timing and duration of recording", *Clin Neurophysiol*, vol. 130, no. 11, pp. 2060–2064, 2019.

[69] C. Baumgartner, J. P. Koren, and M. Rothmayer, "Automatic computer-based detection of epileptic seizures", *Frontiers in Neurology*, vol. 9, p. 639, 2018.

[70] U. R. Acharya, S. V. Sree, G. Swapna, R. J. Martis, and J. S. Suri, "Automated EEG analysis of epilepsy: A review", *Knowledge-Based Systems*, vol. 45, pp. 147–165, 2013.

[71] T. W. Kjaer, H. B. Sorensen, S. Groenborg, C. R. Pedersen, and J. Duun-Henriksen, "Detection of paroxysms in long-term, single-channel EEG-monitoring of patients with typical absence seizures", *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 5, 2017.

[72] P. Xanthopoulos, S. Rebennack, C. C. Liu, J. Zhang, G. L. Holmes, B. M. Uthman, and P. M. Pardalos, "A novel wavelet based algorithm for spike and wave detection in absence epilepsy", in *10th IEEE International Conference on Bioinformatics and Bioengineering 2010, BIBE 2010*, Philadelphia, Pennsylvania, USA: IEEE, 2010, pp. 14–19.

[73] H. Adeli, Z. Zhou, and N. Dadmehr, "Analysis of EEG records in an epileptic patient using wavelet transform", *Journal of Neuroscience Methods*, vol. 123, no. 1, pp. 69–87, 2003.

[74] V. Sakkalis, G. Giannakakis, C. Farmaki, A. Mousas, M. Pediaditis, P. Vorgia, and M. Tsiknakis, "Absence seizure epilepsy detection using linear and nonlinear EEG analysis methods", *Conf Proc IEEE Eng Med Biol Soc*, vol. 2013, pp. 6333–6336, 2013.

[75] K. Zeng, J. Yan, Y. Wang, A. Sik, G. Ouyang, and X. Li, "Automatic detection of absence seizures with compressive sensing EEG", *Neurocomputing*, vol. 171, pp. 497–502, 2016.

[76] F. Manzouri, S. Heller, M. Dümpelmann, P. Woias, and A. Schulze-Bonhage, "A comparison of machine learning classifiers for energy-efficient implementation of seizure detection", *Frontiers in Systems Neuroscience*, vol. 12, p. 43, 2018.

[77] A. Burrello, L. Cavigelli, K. Schindler, L. Benini, and A. Rahimi, "Laelaps: An energy-efficient seizure detection algorithm from long-term human iEEG recordings without false alarms", in *Proceedings of the 2019 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Florence, Italy: IEEE, 2019, pp. 752–757.

[78] M. Hugle, S. Heller, M. Watter, M. Blum, F. Manzouri, M. Dumpelmann, A. Schulze-Bonhage, P. Woias, and J. Boedecker, "Early seizure detection with an energy-efficient convolutional neural network on an implantable microcontroller", *arXiv e-prints*, 2018. arXiv: 1806.04549 [stat.ML].

[79] J. Wouters, F. Kloosterman, and A. Bertrand, "Towards online spike sorting for high-density neural probes using discriminative template matching with suppression of interfering spikes", *Journal of Neural Engineering*, vol. 15, no. 5, 2018.

[80] J. R. Tenney and T. A. Glauser, "The current state of absence epilepsy: can we have your attention?", *Epilepsy Curr*, vol. 13, no. 3, pp. 135–140, 2013.

[81] G. A. Worrell, T. D. Lagerlund, and J. R. Buchhalter, "Role and limitations of routine and ambulatory scalp electroencephalography in diagnosing and managing seizures", *Mayo Clinic Proceedings*, vol. 77, no. 9, pp. 991–998, 2002.

[82] B. D. V. Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering", *IEEE ASSP Magazine*, vol. 5, pp. 4–24, 2 1988.

[83] J. Wouters, F. Kloosterman, and A. Bertrand, "A data-driven regularization approach for template matching in spike sorting with high-density neural probes", in *Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 4376–4379.

[84] S. Beniczky and P. Ryvlin, "Standards for testing and clinical validation of seizure detection devices", *Epilepsia*, vol. 59, no. S1, pp. 9–13, 2018.

[85] R. Hussein, H. Palangi, R. K. Ward, and Z. J. Wang, "Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals", *Clinical Neurophysiology*, vol. 130, no. 1, pp. 25–37, 2019.

[86] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals", *Computers in Biology and Medicine*, vol. 100, pp. 270–278, 2018.

[87] C. Baumgartner and J. P. Koren, "Seizure detection using scalp-EEG", *Epilepsia*, vol. 59, no. S1, pp. 14–22, 2018.

[88] A. M. Narayanan and A. Bertrand, "Analysis of Miniaturization Effects and Channel Selection Strategies for EEG Sensor Networks with Application to Auditory Attention Detection", *IEEE Trans. Biomed. Eng.*, vol. 67, no. 1, pp. 234–244, 2020.

[89] S. G. Johnson and M. Frigo, "Implementing FFTs in practice", in *Fast Fourier Transforms*, C. S. Burrus, Ed., OpenStax CNX, 2012, ch. 11.

[90] M. J. Vrhel, C. Lee, and M. A. Unser, "Comparison of algorithms for the fast computation of the continuous wavelet transform", in *Wavelet Applications in Signal and Image Processing IV*, International Society for Optics and Photonics, vol. 2825, SPIE, 1996, pp. 422–431.

[91] C. Carlson, "Can we screen EEGs more efficiently? spectrographic review of EEG data", *Epilepsy Currents*, vol. 15, p. 24, 1 2015.

[92] Persyst. "Seizure detection and seizure probability". (2022), [Online]. Available: https://www.persyst.com/seizure-detection-and-seizure-probability/ (visited on 10/12/2022).

[93] S. P. Chepuri and G. Leus, "Sparsity-Promoting Sensor Selection for Non-Linear Measurement Models", *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 684–698, 2015.

[94] S. A. Hamza and M. G. Amin, "Sparse Array Beamforming Design for Wideband Signal Models", *IEEE Trans. Aerosp. Electron. Syst.*, vol. 57, no. 2, pp. 1211–1226, 2021.

[95] M. Gao, K. F. C. Yiu, and S. Nordholm, "On the Sparse Beamformer Design", *Sensors*, vol. 18, no. 10, 2018.

[96] W. Shi, Y. Li, L. Zhao, and X. Liu, "Controllable Sparse Antenna Array for Adaptive Beamforming", *IEEE Access*, vol. 7, pp. 6412–6423, 2019.

[97] S. A. Hamza and M. G. Amin, "Optimum sparse array receive beamforming for wideband signal model", in *Proc. of the 52nd ACSSC*, 2018, pp. 89–93.

[98] S. A. Hamza and M. G. Amin, "Sparse Array DFT Beamformers for Wideband Sources", in *Proc. of the IEEE RadarConf19*, 2019, pp. 1–5.

[99] S. A. Hamza and M. G. Amin, "Sparse Array Receiver Beamformer Design for Multi-Functional Antenna", in *Proc. of EUSIPCO 2020*, 2021, pp. 1836–1840.

[100] S. A. Hamza, W. Zhai, X. Wang, and M. G. Amin, "Sparse Array Transceiver Design for Enhanced Adaptive Beamforming in MIMO Radar", in *Proc. of ICASSP 2021*, 2021, pp. 4410–4414.

[101] W. Zhai, X. Wang, S. A. Hamza, and M. G. Amin, "Cognitive-Driven Optimization of Sparse Array Transceiver for MIMO Radar Beamforming", in *Proc. of the IEEE RadarConf21*, 2021, pp. 1–6.

[102] T. Alotaiby, F. E. El-Samie, S. A. Alshebeili, and I. Ahmad, "A review of channel selection algorithms for EEG signal processing", *EURASIP J. Adv. Signal Process.*, no. 66, 2015.

[103] A. M. Narayanan, P. Patrinos, and A. Bertrand, "Optimal Versus Approximate Channel Selection Methods for EEG Decoding With Application to Topology-Constrained Neuro-Sensor Networks", *IEEE Trans. Neural Syst. Rehabilitation Eng.*, vol. 29, pp. 92–102, 2021.

[104] J. Dan, B. Vandendriessche, W. V. Paesschen, D. Weckhuysen, and A. Bertrand, "Computationally-efficient algorithm for real-time absence seizure detection in wearable electroencephalography", *International Journal of Neural Systems*, vol. 30, 11 2020.

[105] A. Bertrand, "Applications and trends in wireless acoustic sensor networks: A signal processing perspective", in *Proc. 18th IEEE SCVT*, 2011, pp. 1–6.

[106] J. Zhang, S. P. Chepuri, R. C. Hendriks, and R. Heusdens, "Microphone Subset Selection for MVDR Beamformer Based Noise Reduction", *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 3, pp. 550–563, 2018.

[107] O. Mehanna, N. D. Sidiropoulos, and G. B. Giannakis, "Joint Multicast Beamforming and Antenna Selection", *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2660–2674, 2013.

[108] S. A. Hamza and M. G. Amin, "Hybrid Sparse Array Beamforming Design for General Rank Signal Models", *IEEE Trans. Signal Process.*, vol. 67, no. 24, pp. 6215–6226, 2019.

[109] S. Geirnaert, T. Francart, and A. Bertrand, "Fast EEG-based decoding of the directional focus of auditory attention using common spatial patterns", *IEEE Trans. Biomed. Eng.*, vol. 68, no. 5, pp. 1557–1568, 2021.

[110] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Muller, "Optimizing spatial filters for robust EEG single-trial analysis", *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, 2007.

[111] M. Dash and H. Liu, "Feature Selection for Classification", *Intell. Data Anal.*, vol. 1, no. 1, pp. 131–156, 1997.

[112] S. Joshi and S. Boyd, "Sensor selection via convex optimization", *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, 2009.

[113] S. Yan and X. Tang, "Trace quotient problems revisited", in *Computer Vision – ECCV 2006*, A. Leonardis, H. Bischof, and A. Pinz, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 232–244.

[114] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite Relaxation of Quadratic Optimization Problems", *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, 2010.

[115] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing Sparsity by Reweighted ℓ_1 Minimization", *J. Fourier Anal. Appl.*, vol. 14, no. 5-6, pp. 877–905, 2008.

[116] M. Grant and S. Boyd, *CVX: Matlab Software for Disciplined Convex Programming, version 2.2*, http://cvxr.com/cvx, 2020.

[117] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs", in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds., http://stanford.edu/~boyd/graph_dcp.html, Springer-Verlag Limited, 2008, pp. 95–110.

[118] MOSEK ApS, *The mosek optimization toolbox for matlab manual. version 9.1.9*, 2019. [Online]. Available: http://docs.mosek.com/9.1/toolbox/index.html.

[119] G. H. Golub and H. A. Van Der Vorst, "Eigenvalue computation in the 20th century", *J. Comput. Appl. Math.*, vol. 123, no. 1-2, pp. 35–65, 2000.

[120] F. Qi, W. Wu, Z. L. Yu, Z. Gu, Z. Wen, T. Yu, and Y. Li, "Spatiotemporal-Filtering-Based Channel Selection for Single-Trial EEG Classification", *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 558–567, 2021.

[121] J. Meng, G. Liu, G. Huang, and X. Zhu, "Automated selecting subset of channels based on CSP in motor imagery brain-computer interface system", in *Proc. of IEEE Int. Conf. ROBIO*, 2009, pp. 2290–2294.

[122] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Optimizing the Channel Selection and Classification Accuracy in EEG-Based BCI", *IEEE Trans. Biomed. Eng.*, vol. 58, no. 6, pp. 1865–1873, 2011.

[123] I. Onaran, N. F. Ince, and A. E. Cetin, "Sparse spatial filter via a novel objective function minimization with smooth ℓ_1 regularization", *Biomed. Signal Process. Control*, vol. 8, no. 3, pp. 282–288, 2013.

[124] A. Galecki and T. Burzykowski, *Linear Mixed-Effects Models Using R: A Step-by-Step Approach* (Springer Texts in Statistics). Springer-Verlag New York, 2013.

[125] B. U. Westner, S. S. Dalal, A. Gramfort, V. Litvak, J. C. Mosher, R. Oostenveld, and J.-M. Schoffelen, "A unified view on beamformers for M/EEG source reconstruction", *NeuroImage*, vol. 246, p. 118 789, 2022.

[126] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices", *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365–411, 2004.

[127] M. O. Baud, J. K. Kleen, E. A. Mirro, J. C. Andrechak, D. King-Stephens, E. F. Chang, and V. R. Rao, "Multi-day rhythms modulate seizure risk in epilepsy", *Nature Communications*, vol. 9, pp. 1–10, 1 2018.

[128] P. J. Karoly, D. R. Freestone, R. Boston, D. B. Grayden, D. Himes, K. Leyde, U. Seneviratne, S. Berkovic, T. O'Brien, and M. J. Cook, "Interictal spikes and epileptic seizures: Their relationship and underlying rhythmicity", *Brain*, vol. 139, pp. 1066–1078, 4 2016.

[129] R. Krendl, S. Lurger, and C. Baumgartner, "Absolute spike frequency predicts surgical outcome in tle with unilateral hippocampal atrophy", *Neurology*, vol. 71, pp. 413–418, 6 2008.

[130] J. Duun-Henriksen, T. W. Kjaer, R. E. Madsen, L. S. Remvig, C. E. Thomsen, and H. B. D. Sorensen, "Channel selection for automatic seizure detection", *Clinical Neurophysiology*, vol. 123, pp. 84–92, 1 2012.

[131] K. Vandecasteele, T. D. Cooman, J. Dan, E. Cleeren, S. V. Huffel, B. Hunyadi, and W. V. Paesschen, "Visual seizure annotation and automated seizure detection using behind-the-ear electroencephalographic channels", *Epilepsia*, vol. 61, pp. 766–775, 4 2020.

[132] A. J. Casson, "Wearable eeg and beyond", *Biomedical Engineering Letters*, vol. 9, pp. 53–71, 1 2019.

[133] O. Valentin, G. Viallet, A. Delnavaz, G. Cretot-Richert, M. Ducharme, H. Monsarat-Chanon, and J. Voix, "Custom-fitted in- and around-the-ear sensors for unobtrusive and on-the-go EEG acquisitions: Development and validation", *Sensors*, vol. 21, p. 2953, 9 2021.

[134] M. A. Frankel, M. J. Lehmkuhle, M. C. Spitz, B. J. Newman, S. V. Richards, and A. M. Arain, "Wearable reduced-channel EEG system for remote seizure monitoring", *Frontiers in Neurology*, vol. 12, p. 1842, 2021.

[135] I. C. Zibrandtsen, P. Kidmose, C. B. Christensen, and T. W. Kjaer, "Ear-EEG detects ictal and interictal abnormalities in focal and generalised epilepsy – a comparison with scalp EEG monitoring", *Clinical Neurophysiology*, vol. 128, pp. 2454–2461, 12 2017.

[136] C. M. Epstein and G. P. Brickley, "Interelectrode distance and amplitude of the scalp EEG", *Electroencephalography and Clinical Neurophysiology*, vol. 60, pp. 287–292, 4 1985.

[137] R. Quon, S. Meisenhelter, E. Camp, M. Testorf, Y. Song, Q. Song, G. Culler, P. Moein, and B. Jobst, "Aied: Artificial intelligence for the detection of intracranial interictal epileptiform discharges", *Clinical neurophysiology*, vol. 133, pp. 1–8, 2022.

[138] J. Jin, J. Dauwels, S. Cash, and M. Westover, "Spikegui: Software for rapid interictal discharge annotation via template matching and online machine learning", *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2014, pp. 4435–4438, 2014.

[139] J. Jing, J. Dauwels, T. Rakthanmanon, E. Keogh, S. Cash, and M. Westover, "Rapid annotation of interictal epileptiform discharges via template matching under dynamic time warping", *Journal of neuroscience methods*, vol. 274, pp. 179–190, 2016.

[140] S. Lodder, J. Askamp, and M. van Putten, "Inter-ictal spike detection using a database of smart templates", *Clinical neurophysiology*, vol. 124, pp. 2328–2335, 12 2013.

[141] N. M. Harid, J. Jing, J. Hogan, *et al.*, "Measuring expertise in identifying interictal epileptiform discharges", *Epileptic Disorders*, vol. 24, pp. 496–506, 3 2022.

[142] J. Jing, A. Herlopian, I. Karakis, *et al.*, "Interrater reliability of experts in identifying interictal epileptiform discharges in electroencephalograms", *JAMA neurology*, vol. 77, pp. 49–57, 1 2020.

[143] E. Reus, F. Cox, J. van Dijk, and G. Visser, "Automated spike detection: Which software package?", *Seizure*, vol. 95, pp. 33–37, 2022.

[144] K. M. Bushby, T. Cole, J. N. Matthews, and J. A. Goodship, "Centiles for adult head circumference.", *Archives of Disease in Childhood*, vol. 67, pp. 1286–1287, 10 1992.

[145] G. V. Brummelen, *Heavenly mathematics: The forgotten art of spherical trigonometry*. Princeton University Press, 2012.

[146] S. Steinarsson, "Downsampling time series for visual representation", University of Iceland, 2013.

[147] S. Basu, M. Banerjee, and A. Sen, "Bayesian inference for kappa from single and multiple studies", *Biometrics*, vol. 56, pp. 577–582, 2 2000.

[148] S. Ferrari and F. Cribari-Neto, "Beta regression for modelling rates and proportions", *Journal of Applied Statistics*, vol. 31, pp. 799–815, 7 2004.

[149] J. K. Kruschke, "Bayesian estimation supersedes the t test", *Journal of experimental psychology*, vol. 142, pp. 573–603, 2 2013.

[150] J. Salvatier, T. V. Wiecki, and C. Fonnesbeck, "Probabilistic programming in python using pymc3", *PeerJ Computer Science*, vol. 2016, e55, 4 2016.

[151] J. J. Halford, D. Sabau, F. W. Drislane, T. N. Tsuchida, and S. R. Sinha, "American clinical neurophysiology society guideline 4: Recording clinical eeg on digital media", *Journal of clinical neurophysiology*, vol. 33, pp. 317–319, 4 2016.

[152] *ADS1299-x low-noise, 4-, 6-, 8-channel, 24-bit, analog-to-digital converter for EEG and biopotential measurements datasheet*, ADS1299, Rev. C, Texas Instruments, 2017. [Online]. Available: https://www.ti.com/lit/gpn/ads1299.

[153] M. G. Bleichner, M. Lundbeck, M. Selisky, F. Minow, M. Jäger, R. Emkes, S. Debener, and M. D. Vos, "Exploring miniaturized eeg electrodes for brain-computer interfaces. an eeg you do not see?", *Physiological reports*, vol. 3, 4 2015.

[154] J. Wouters, A. Bertrand, and F. Kloosterman, "Design and validation of low-complexity methods for resolving spike overlap in neuronal spike sorting", Ph.D. dissertation, KU Leuven, 2020.

[155] J. Y. Liou, E. H. Smith, L. M. Bateman, S. L. Bruce, G. M. McKhann, R. R. Goodman, R. G. Emerson, C. A. Schevon, and L. F. Abbott, "A model for focal seizure onset, propagation, evolution, and progression", *eLife*, vol. 9, 2020.

[156] D. P. Dash, M. H. Kolekar, and K. Jha, "Multi-channel EEG based automatic epileptic seizure detection using iterative filtering decomposition and hidden markov model", *Computers in biology and medicine*, vol. 116, 2020.

[157] A. Bertrand and M. Moonen, "Distributed adaptive generalised eigenvector estimation of a sensor signal covariance matrix pair in a fully connected sensor network", *Signal Processing*, vol. 106, pp. 209–214, 2015.

[158] U. Z. Leuven. "Sensor-Dot and Plug 'n Patch Study in Epilepsy". ClinicalTrials.gov, Ed. (2020), [Online]. Available: https://clinicaltrials.gov/ct2/show/NCT04642105 (visited on 10/06/2022).

[159] F. S. Leijten, J. van Andel, C. Ungureanu, *et al.*, "Multimodal seizure detection: A review", *Epilepsia*, vol. 59 Suppl 1, pp. 42–47, 2018.

[160] F. Fürbass, S. Kampusch, E. Kaniusas, J. Koren, S. Pirker, R. Hopfengärtner, H. Stefan, T. Kluge, and C. Baumgartner, "Automatic multimodal detection for long-term seizure documentation in epilepsy", *Clinical neurophysiology*, vol. 128, pp. 1466–1472, 8 2017.

Acknowledgements

This work was carried out at the ESAT-STADIUS Center for Dynamical Systems, Signal Processing, and Data Analytics (Department of Electrical Engineering), KU Leuven and at Byteflies.

Jonathan Dan was supported by a Baekeland mandate from Flanders Innovation & Entrepreneurship (VLAIO) (no. HBC.2018.0189). Furthermore, the work in this thesis was supported by FWO project no. G0A4918N, the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (grant agreement no. 802895 and no. 637424), and the Flemish Government (AI Research Program).

Biography

Jonathan Dan was born in Brussels (Belgium) on 24 May 1992. He studied engineering at the Université libre de Bruxelles. He had his first experience with EEG during a bachelor project studying the feasibility of a low-cost EEG platform for use in rural Rwanda. The project was done in collaboration with the University Hospital of Kigali and local partners. It was presented at a meeting of the European Society of Paediatric Neurology. He then completed his first year of masters at TU Delft through an Erasmus exchange programme where he studied biomedical electronics. His first direct involvement in academic research was during an internship at Prof. Jean Gotman's lab at The Neuro Institute at McGill university, where he studied the relationship between scalp and intracranial EEG. This led to his first journal publication, in NeuroImage. After graduating with a MSc degree in Biomedical Engineering (magna cum laude) from ULB in 2016, he joined Byteflies in Antwerp, where he contributed to the development of wearable sensors and to SeizeIT, a collaboration between Byteflies, KU Leuven, UZ Leuven, Pilipili and UCB, exploring sensing modalities for home monitoring of people with epilepsy. When this project was completed, he successfully applied for a Baekeland mandate, which allowed him to do this PhD in a partnership between ESAT (KU Leuven) and Byteflies.

Jonathan was finalist of the Prologin French National Computer Science Contest in 2012. He received a best poster award at the Belgian National Day on Biomedical Engineering in 2019. With ESAT colleagues, he won the Neureka epilepsy detection challenge in 2020 and received the best paper award at the 2020 IEEE Signal Processing in Medicine and Biology Symposium. During his PhD, he was active in a number of science outreach activities, and his research has been covered by various national and international media.

List of publications

5.3 International Journal Publications

- **Dan J**, M.T. Foged M.T, Vandendriessche B, Van Paesschen W, Bertrand A. Sensor selection and miniaturization limits for detection of interictal epileptiform discharges with wearable EEG. *Under review*
- Pelc K, Gajewska A, Napiórkowski N, **Dan J**, Verhoeven C, Dan B. Multiscale entropy as a metric of brain maturation in a large cohort of typically developing children born preterm using longitudinal high-density EEG in the first two years of life. *Physiological Measurement (2022)*
- **Dan J**, Geirnaert S, Bertrand A. Grouped Variable Selection for Generalised Eigenvalue Problems. *Elsevier Signal Processing (2022), Volume 195 (108476)*
- **Dan J**. Wearable sensors to improve activities in individuals with cerebral palsy. *Developmental medicine and child neurology (2020), 62(6), 664*
- Vandecasteele K, De Cooman T, **Dan J**, Cleeren E, Van Huffel S, Hunyadi B, Van Paesschen W. Visual seizure annotation and automated seizure detection using behind-the-ear electroencephalographic channels. *Epilepsia (2020), 61(4), 766–775*
- **Dan J**, Vandendriessche B, Van Paesschen W, Weckhuysen D, Bertrand A. Computationally-Efficient Algorithm for Real-Time Absence Seizure Detection in Wearable Electroencephalography. *International Journal of Neural Systems (2020), 30(11)*
- Gu Y, Cleeren E, **Dan J**, Claes K, Van Paesschen W, Van Huffel S, Hunyadi B. Comparison between Scalp EEG and Behind-the-Ear EEG for Development of a Wearable Seizure Detection System for Patients with Focal Epilepsy. *Sensors (2018), 18(1)*

5.4 Conference Proceedings

Chatzichristos C, **Dan J**, Narayanan AM, Seeuws N, Vandecasteele K, De Vos M, Bertrand A, Van Huffel S. Epileptic Seizure Detection in EEG via Fusion of Multi-View Attention-Gated U-net Deep Neural Networks. *IEEE signal processing in medicine and biology symposium (IEEE SPMB 2020)*. Winner of the Neureka Challenge and best paper award at the IEEE SPMB 2020 symposium

5.5 Conference abstracts and oral communications

Dan J, Vandendriessche B, Van Paesschen W, Bertrand A. Optimal electrode selection for mobile EEG. Oral presentation at the *2021 Mobile Devices and Seizure Detection in Epilepsy*

Dan J, Vandendriessche B, Van Paesschen W, Weckhuysen D, Bertrand A. Low-Complexity algorithm for absence seizure detection in wearable electroencephalography. Poster presented at the *2019 Belgian National Day on Biomedical Engineering*. Best poster award (3^{rd})

Dan J, Vandendriessche B, Van Paesschen W, Weckhuysen D, Bertrand A. Novel algorithm for seizure detection in wearable electroencephalography. Oral presentation at the *2019 Mobile Devices and Seizure Detection in Epilepsy*

Dan J, Vandendriessche B, Van Paesschen W, Weckhuysen D, Bertrand A. Low-complexity data-driven seizure detection algorithm for home monitoring of patients with epilepsy using wearable EEG. Poster presentation at the *41st Annual International Conference of the IEEE Engineering in Medicine & Biology Society*

Dan J, Weckhuysen D, Cleeren E, Van Paesschen W, Vandendriessche B. Technical Validation of Sensor Dot: A Wearable for Ambulatory Monitoring of Epileptic Seizures. Poster presentation at the *41st Annual International Conference of the IEEE Engineering in Medicine & Biology Society*

FACULTY OF ENGINEERING SCIENCE
DEPARTMENT OF ELECTRICAL ENGINEERING
BIOMED
Celestijnenlaan 200A box 2402
B-3001 Leuven
jonathan.dan@kuleuven.be
https://www.esat.kuleuven.be/stadius/