Czech Technical University in Prague
Faculty of Electrical Engineering
Department of Cybernetics

Automatic Sleep EEG Pattern Detection

Doctoral Thesis

Mgr. Elizaveta Saifutdinova

Ph.D. Study Programme: P2612 - Electrical Engineering and Information Technology
Branch of Study: 3902V035 - Artificial Intelligence and Biocybernetics
Supervisor: **Doc. Ing. Lenka Lhotska, CSc.**
Supervisor-Specialist: **Ing. Vaclav Gerla, Ph.D.**

Prague, June 2019

Supervisor:
Doc. Ing. Lenka Lhotska, CSc.
Department of Natural Sciences
Faculty of Biomedical Engineering
and Czech Institute of Informatics, Robotics, and Cybernetics
Czech Technical University in Prague
Jugoslávskych partyzanů 1580/3
160 00 Prague 6
Czech Republic

Supervisor-Specialist:
Ing. Vaclav Gerla, Ph.D.
Czech Institute of Informatics, Robotics, and Cybernetics
Czech Technical University in Prague
Jugoslávskych partyzanů 1580/3
160 00 Prague 6
Czech Republic

Declaration

I hereby declare I have written this doctoral thesis independently and quoted all the sources of information used in accordance with methodological instructions on ethical principles for writing an academic thesis. Moreover, I state that this thesis has neither been submitted nor accepted for any other degree.

In Prague, June 2019

.....
Mgr. Elizaveta Saifutdinova

Abstract

Analysis of recorded brain activity is one of the main investigation methods in modern sleep medicine and research. Long duration and complex nature of the data make it difficult for manual investigation. Moreover, high inter-subject variability could cause problems in automatic processing. This thesis focuses on the problem of automatic EEG pattern detection. We concentrate on artifacts and sleep spindles as two typical patterns in sleep EEG. We review the methodologies and strategies used in real sleep research practice. Moreover, we investigate and test the performance of state-of-the-art approaches for the tasks. We propose enhancement methodologies and use expert's strategies for automatic method development. The proposed methods utilize recent advances in EEG pattern detection. They are adaptive and fully unsupervised. In the thesis, testing is performed on the data collected from subjects suffering from a sleep disorder which increases inter-subject variability of the data. We analyze obtained results in aspects of formal statistical measures and using visual inspection of the data providing more details about data nature.

Keywords: sleep EEG, pattern detection, machine learning, clustering, adaptive methods, Riemannian geometry

Abstrakt

Analýza mozkové aktivity je jednou z klíčových vyšetřovacích metod v moderní spánkové medicíně a výzkumu. Ruční vyhodnocování EEG záznamů komplikuje rozmanitost a délka naměřených signálů. Zejména vysoká interpersonální variabilita způsobuje problémy při automatickém zpracování. Hlavním cílem disertační práce je automatická detekce EEG grafoelementů. Navržená metodologie umožnuje rozpoznávat klinicky významné spánkové vzory jako artefakty a spánková vřeténka. V rámci této práce byly zkoumány různé metodologie využívané výzkumnými spánkovými laboratořemi a otestována výkonnost nejmodernějších přístupů v této oblasti. Byla navržena metodologie, umožňující integraci efektivnějších metod rozpoznávání a využití expertních strategií pro vývoj automatických metod pro hodnocení spánkových dat. Využity byly nejnovější poznatky v oblasti detekce EEG vzorů, zejména v oblasti adaptivního zpracování signálu a metod učení bez učitele. Metody byly použity na reálné klinické záznamy subjektů se spánkovou poruchou, u nichž je velmi vysoká interpersonální variabilita EEG dat. Získané výsledky byly statisticky validovány. Nedílnou součástí práce je visuální inspekce výstupů v rámci celé navržené metodologie, což umožní lépe interpretovat získané výsledky a získat jasnější představu o povaze analyzovaných dat.

Klíčová slova: spánkové EEG, detekce grafoelementů, strojové učení, shlukování, adaptivní methody, Riemannová geometrie

Acknowledgements

I am deeply grateful to my supervisor, Doc. Ing. Lenka Lhotska, CSc for her thoughtful guidance and, no less important, for believing in me and giving me that amazing opportunity to become a part of a scientific society. I would particularly like to thank my husband Dimitry for his support and endless cookie supply and my dearest friend Evgenia Youett, Ph.D. for inspiration for beginning this path. Also, I would like to thank all the people who helped me, challenged me and shaped my thoughts along the whole path: my co-supervisor Vaclav Gerla, my colleagues from Cog Sys CIIRC Martin Macas, Iva Novotna, Mira Bursa, Jirka Spilka, Vaclav Chudacek and Michal Huptych; my colleagues from VP5 NIMH Daniela Dudysova, Jana Koprivova, Eva Farkova, Michal Smotek, Karolina Veldova and Monika Klikova; my co-author Marco Congedo.

Research has been partially supported by "Temporal context in analysis of long-term non-stationary multidimensional signal", register number 17-20480S of the Grant Agency of the Czech Republic and by the project nr. LO1611 with a financial support from the MEYS under the NPU I program.

List of Tables

List of Figures

List of Acronyms and Abbreviations

AASM American Academy of Sleep Medicine. 9

AFAST proposed adaptive artifact detection method based on FASTER features. 34

BCI brain-computer interface. 45

BSS blind source separation. 16

DBSCAN density-based spatial clustering of applications with noise. 27

ECG electrocardiogram. 10

EEG electroencephalogram. 7

EM expectation maximization. 26

EMG electromyogram. 10

EOG electrooculogram. 10

FASST artifact detection strategy proposed for fMRI artefact rejection and sleep scoring toolbox. 37

FASTER fully automated, unsupervised method for processing of high density EEG data. 17

FN false negative. 28

FP false positive. 28

GMM Gaussian mixture model. 25

ICA independent component analysis. 16

ICSD3 International Classification of Sleep Disorders third edition. 13

non-REM non-rapid eye movement. 11

PLM periodic limb movement. 15

PSG polysomnogram. 9

RBD REM sleep behavior disorder. 14

REM rapid eye movement. 11

RMS sleep spindle detector based on root mean square. 19

RP Riemannian potato artifact detector. 48

RPs Riemannian potatoes artifact detection method. 49

RSP sleep spindle detector using a relative sigma power. 19

SCADM artifact detection strategy based on single channel artifact detection methods. 37

SCM sample covariance matrix. 47

SIGMA sleep spindle detector utilizing S-transform and defining sigma index computed for obtained frequencies. 19

SPD symmetric positive-definite. 47

SSAD proposed sleep spindle adaptive detector. 63

std standard deviation. 24

SVM support vector machine. 19

TEO sleep spindle detector based on Teager energy operator. 19

TN true negative. 28

TNR true negative rate. 29

TP true positive. 28

TPR true positive rate. 29

Contents

Chapter 1

Introduction

Analysis of recorded electrical brain activity (EEG, see in Section 2.1.2) is a widely used method for brain investigation. Its inexpensiveness and low technical requirements caused spreading it into many research fields and, in particular, in sleep medicine. Sleep is an unconscious process, objective observation and measurements are required for conducting research in this field. Recorded brain activity along with other body measurements (PSG, see in Section 2.1.2) serve as a source of information about internal physiological processes. Specific processes can be recognized from signal patterns occurring in sleep recordings. Analysis of such signal patterns makes a great impact on sleep research by providing the connection between pattern properties and other physiological/psychological traits of the investigated subject.

The thesis focuses on the problem of automatic identification of patterns in sleep EEG. Long duration, variation of patterns and their simultaneous occurrence make it difficult for manual investigation. The main objective of the study is the design and development of methods for automatic identification of EEG patterns, which could perform well on sleep data. It also should be stable to inter-subject variability of data, which is even more severe for data collected from subjects suffering from sleep disorders. To overcome this problem, we propose methods based on adaptive techniques and unsupervised machine learning approaches which allow for signal processing independent of examples collected from other subjects. In this study, we concentrate on brain signals only; however, the proposed methods could be easily extended to other channels of sleep recording.

Detection of two kinds of EEG patterns is discussed in the study. The first one is an artifact pattern, which can be defined only by comparison to normal EEG data. This is a typical example of outlier detection. It is discussed in two aspects: a detailed artifact detection for short-term sleep data and approach for processing a whole night recording. The proposed methods based on machine learning are discussed regarding state-of-the-art methods. The other typical kind of EEG pattern is described by spectral and amplitude

characteristics. In this case, we concentrate on sleep spindle detection in sleep EEG. We propose a new adaptive method and compare it to a traditional approach often used in clinical practice and neurological research. Finally, we examine how to combine results of previously developed methods. Using detection from several methods at once instead of choosing only one detector benefits for visual data inspection.

In the thesis, we focus on the real problems that occurred in a sleep research laboratory. Testing is concentrated on processing data recorded from patients suffering from insomnia and other sleep disorders due to high variability in terms of similarity of data characteristics in comparison to recordings of healthy subjects. All methods are implemented and used for data research and data processing in the sleep laboratory in the National Institute of Mental Health, Czech Republic. Implementation details of the methods are provided as well.

1.1 Goals of the thesis

The main goal of the thesis is to propose adaptive and effective methods for automatic sleep EEG pattern detection. The methods should be adaptive to overcome inter-subject variability caused by sleep disorders. In the described research, we investigate problems of detection of two types of EEG patterns: artifacts and sleep spindles. The main goals may be summarized as follows:

- To investigate the problem of automated pattern detection in a sleep EEG recorded as a part of a whole night PSG. To review a pipeline for sleep PSG/EEG data processing and pattern detection used in real practice. To test state-of-the-art-methods on data collected from patients suffering from a sleep disorder.

- To study the problem of artifact detection in a long-term EEG for a practical application. To propose and test an adaptive and unsupervised improvement of the existing method to increase adaptability.

- To research spectral properties of sleep EEG and propose an automatic abnormality detection using Riemannian geometry. To implement a designed solution and test it on real sleep EEG recorded from subjects suffering from sleep disorders.

- To study the problem of adaptive segmentation for automatic pattern detection with a focus on sleep spindle detection as a domain problem. To propose an adaptive sleep spindle detection method and test it on the real sleep EEG data from subjects suffering from sleep disorders.

- To investigate the problem of aggregation of outcomes of existing methods for detection visualization. To investigate the strategy used by experts and implement a solution, which provides supporting information to manual data investigation.

1.2 Thesis organization

The rest of the thesis is organized as follows. Chapter 2 provides medical and technical background for brain signal measurements, sleep recordings, and automatic pattern detection in biomedical signals. An artifact detection pipeline for processing of long-term EEG and an proposed artifact detection method is presented in Chapter 3. Next, Chapter 4 concentrates on artifact detection in a short-term sleep EEG and presents a novel artifact detection method based on Riemannian geometry. Chapter 5 focuses on the problem of sleep spindle detection in one channel EEG and describes a novel proposed adaptive method. The problem of aggregation of results obtained by previously developed methods is investigated in Chapter 6. The last Chapter 7 summarizes the thesis and provides an overview of succeeding in achievements of goals stated in Section 1.1.

1.3 List of Author's Publication

Journals with impact factor

- E. Saifutdinova, M. Congedo, D. Dudysova, L. Lhotska, J. Koprivova, and V. Gerla, "An unsupervised multichannel artifact detection method for sleep eeg based on riemannian geometry", *Sensors*, vol. 19, no. 3, 2019, ISSN: 1424-8220. DOI: 10.3390/s19030602 (16/16/16/16/16/16)

- V. Gerla, V. Kremen, N. Covassin, L. Lhotska, E. Saifutdinova, J. Bukartyk, V. Marik, and V. Somers, "Automatic identification of artifacts and unwanted physiologic signals in eeg and eog during wakefulness", *Biomedical Signal Processing and Control*, vol. 31, pp. 381 –390, 2017, ISSN: 1746-8094. DOI: 10.1016/j.bspc.2016.09.006 (12/12/12/12/12/12/12/12)

- D. Dudysova, K. Veldova, M. Smotek, E. Saifutdinova, J. Koprivova, J. Buskova, B. A. Mander, M. Brunovsky, P. Zach, J. Korcak, V. Andrashko, M. Viktorinova, F. Tyls, A. Bravermanova, T. Froese, T. Palenicek, and J. Horacek, "Effects of daytime administration of psilocybin on sleep: Similar changes to antidepressants?", *Frontiers in Pharmacology*, 2019 (submitted)

Conferences

Publications are listed in chronological order (from newest to oldest).

- E. Saifutdinova, D. U. Dudysová, V. Gerla, and L. Lhotská, "Improvement sleep spindle detection by aggregation techniques", in *15th Mediterranean Conference on Medical and Biological Engineering and Computing (MEDICON2019)*, 2019 (accepted) (25/25/25/25)

- E. Saifutdinova, D. U. Dudysová, L. Lhotská, V. Gerla, and M. Macaš, "Artifact detection in multichannel sleep eeg using random forest classifier", in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Dec. 2018, pp. 2803–2805. DOI: 10.1109/BIBM.2018.8621374 (20/20/20/20/20)

- M. Macas, N. Grimova, V. Gerla, L. Lhotska, and E. Saifutdinova, "Active learning for semiautomatic sleep staging and transitional eeg segments", in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Dec. 2018, pp. 2621–2627. DOI: 10.1109/BIBM.2018.8621339 (20/20/20/20/20)

- V. Gerla, V. Kremen, M. Macas, E. Saifutdinova, A. Mladek, and L. Lhotska, "Expert-in-the-loop learning for sleep eeg data", in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Dec. 2018, pp. 2590–2596. DOI: 10.1109/BIBM.2018.8621557 (16/16/16/16/16/16)

- V. Gerla, M. Murgas, A. Mladek, E. Saifutdinova, M. Macas, and L. Lhotska, "Hybrid hierarchical clustering algorithm used for large datasets: A pilot study on long-term sleep data", in *Precision Medicine Powered by pHealth and Connected Health*, N. Maglaveras, I. Chouvarda, and P. de Carvalho, Eds., Singapore: Springer Singapore, 2018, pp. 3–7, ISBN: 978-981-10-7419-6 (16/16/16/16/16/16)

- V. Gerla, E. Saifutdinova, M. Macas, A. Mladek, and L. Lhotska, "P01-comparison of short-time fourier transform and continuous wavelet transform for frequency analysis of sleep eeg", *Clinical Neurophysiology*, vol. 129, no. 4, e14, 2018, ISSN: 1388-2457. DOI: 10.1016/j.clinph.2018.01.046 (20/20/20/20/20)

- E. Saifutdinova, V. Gerla, M. Macas, and L. Lhotska, "P04-spatial geometric analysis in sleep polysomnographic data", *Clinical Neurophysiology*, vol. 129, no. 4, e15, 2018, ISSN: 1388-2457. DOI: 10.1016/j.clinph.2018.01.049 (25/25/25/25)

- E. Saifutdinova, V. Gerla, and L. Lhotská, "Riemannian geometry in sleep stage classification", in *Information Technology in Bio- and Medical Informatics*, M. Bursa,

A. Holzinger, M. E. Renda, and S. Khuri, Eds., Cham: Springer International Publishing, 2017, pp. 92–99, ISBN: 978-3-319-64265-9 (33/33/33)

- E. Saifutdinova, V. Gerla, M. Macaš, and L. Lhotská, "Spatial geometric analysis in sleep polysomnographic data", in *64. SPOLEČNÝ SJEZD ČESKÉ A SLOVENSKÉ SPOLEČNOSTI PRO KLINICKOU NEUROFYZIOLOGII.*, Sep. 2017, pp. 26–27 (25/25/25/25)

- V. Gerla, E. Saifutdinova, M. Macaš, A. Mládek, and L. Lhotská, "Comparison of short-time fourier transform and continuous wavelet transform for frequency analysis of sleep eeg", in *64. SPOLEČNÝ SJEZD ČESKÉ A SLOVENSKÉ SPOLEČNOSTI PRO KLINICKOU NEUROFYZIOLOGII.*, Sep. 2017, pp. 26–27 (20/20/20/20/20)

- K. Veldová, D. U. Dudysová, E. Saifutdinová, and J. Kopřivová, "Comparison of sleep characteristics in chronic insomnia subtypes: Eeg correlates of sleep misperception", *Sleep Medicine*, vol. 40, e337 –e338, 2017, ISSN: 1389-9457. DOI: `10.1016/j.sleep.2017.11.992` (25/25/25/25)

- M. Šmotek, P. Vlček, E. Saifutdinova, and J. Kopřivová, "Does blue-light increase cortical activation? findings from an eeg study", *Sleep Medicine*, vol. 40, e310 –e311, 2017, ISSN: 1389-9457. DOI: `10.1016/j.sleep.2017.11.914` (25/25/25/25)

- D. Dudysová, I. Malá, K. Mladá, E. Saifutdinova, J. Koprivova, and P. Sos, "Structural and construct validity of the czech version of the pittsburgh sleep quality index in chronic insomnia", *Neuro endocrinology letters*, vol. 38, no. 1, 67—73, Feb. 2017, ISSN: 0172-780X (16/16/16/16/16/16)

- E. Saifutdinova, M. Macaš, V. Gerla, and L. Lhotská, "Adaptive segmentation optimization for sleep spindle detector", in *Information Technology in Bio- and Medical Informatics*, M. E. Renda, M. Bursa, A. Holzinger, and S. Khuri, Eds., Cham: Springer International Publishing, 2016, pp. 85–96, ISBN: 978-3-319-43949-5 (25/25/25/25)

- E. Saifutdinova, J. Koprivova, L. Lhotska, and M. Macas, "Topological properties of functional brain connectivity in obsessive-compulsive disorder", in *XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016*, E. Kyriacou, S. Christofides, and C. S. Pattichis, Eds., Cham: Springer International Publishing, 2016, pp. 157–161, ISBN: 978-3-319-32703-7 (25/25/25/25)

- E. Saifutdinova, V. Gerla, L. Lhotska, J. Koprivova, and P. Sos, "Sleep spindles detection using empirical mode decomposition", in *Computational Intelligence for*

Multimedia Understanding (IWCIM), 2015 International Workshop on, Oct. 2015, pp. 1–5 (20/20/20/20/20)

- V. Gerla, E. A. Saifutdinova, V. Kremen, M. Huptych, V. Krajca, and L. Lhotska, "A fully unsupervised clustering on adaptively segmented long-term eeg data", in *World Congress on Medical Physics and Biomedical Engineering, June 7-12, 2015, Toronto, Canada*, D. A. Jaffray, Ed., Cham: Springer International Publishing, 2015, pp. 1002–1005, ISBN: 978-3-319-19387-8 (16/16/16/16/16/16)

- E. Saifutdinova, V. Gerla, and L. Lhotska, "Adaptive segmentation with successive windows", in *Proceedings of BioDat 2014 - Conference on Advanced Methods of Biological Data and Signal Processing*, 2014 (33/33/33)

Chapter 2

Background Knowledge

2.1 Measurements

2.1.1 Electroencephalogram

This section provides biological and technical background on the recording of brain activity. Neurons are the basic data processing units of the brain. They are very closely interconnected via axons and dendrites [79] and communicate with each other by sending electrical impulses, which arrive simultaneously and, then, are added together. It leads to the generation of an electrical discharge, known as action potential (a “nerve impulse"). The action potential forms the input to the next neuron in the network. The goal of electroencephalography monitoring method is to record this electrical activity. Obtained data are called electroencephalogram (EEG).

Recordings made from electrodes placed on the surface of the scalp using a distance reference electrode are the resultant field potential at the boundary of a large volume conductor containing many active neurons. Action potentials in axons contribute little to scalp surface records as they are asynchronous and the axons run in many different directions. Surface records are the net effect of local postsynaptic potentials of cortical cells. These may be both excitatory and inhibitory [79]. The internationally standardized 10-20 system is employed for routine clinical practice. In this system, electrodes are located on the surface of the scalp by specific measurements taken between constant anatomical landmarks to determine the placement of electrodes. Exactly, the distance from the nasion (bridge of the nose) to the inion (the lowest point of the skull from the back of the head), and the distance of the preauricular-to-preauricular line as shown in Figure 2.1 [79]. Bipolar or unipolar electrodes can be used in the EEG measurement. In the former method, the potential difference between a pair of electrodes is measured. In the latter method the potential of each electrode is compared either to a neutral electrode

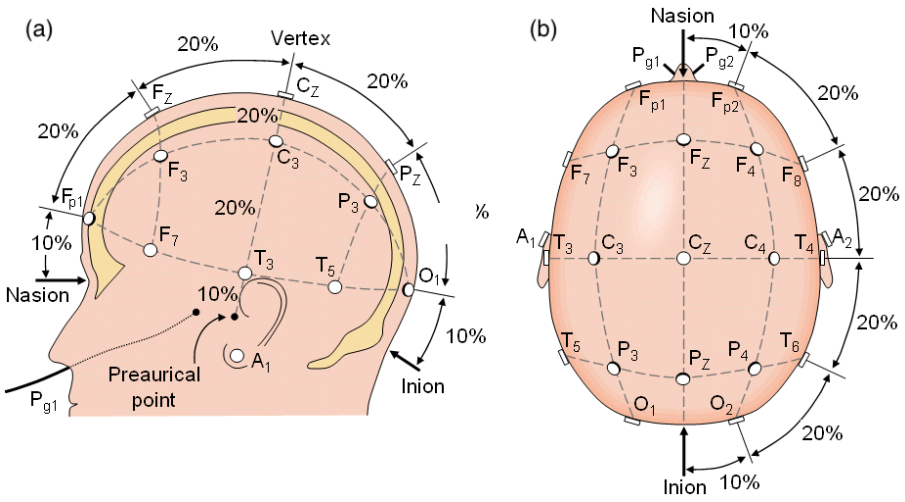


Figure 2.1: The international 10-20 system seen from (**a**) left and (**b**) above the head [79].

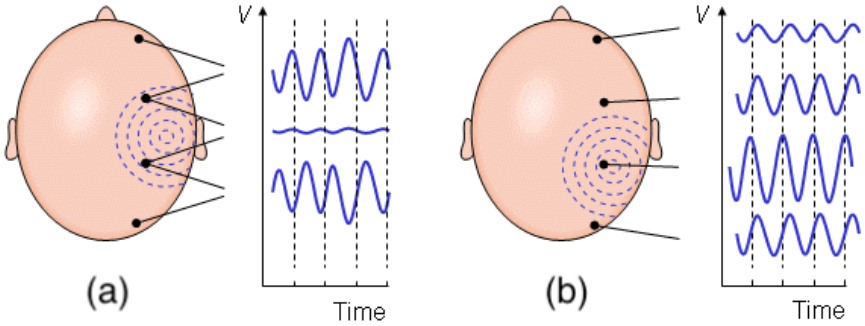


Figure 2.2: Bipolar (**a**) and (**b**) unipolar measurements [79].

or to the average of all electrodes as it is shown in Figure 2.2. Typical EEG waveform depends on the measurement location and reference method.

The basic EEG types can be distinguished based on purpose and methodology of [53], [79]:

- Spontaneous EEG, measured activity that goes on continuously in the living individual over a period of time. The amplitude of the EEG is about 100 μV when measured on the scalp, and about 1–2 μV when measured on the surface of the brain. Frequency is usually under 50 Hz.

- Evoked potentials, components of the EEG that arise in response to a stimulus (which may be electric, auditory, visual, etc.). Such signals are usually below the noise level and thus are not easily distinguished, and one may use signal averaging to improve the signal-to-noise ratio.

- Single-neuron behavior can be examined using microelectrodes, which impale the cells of interest. Through studies of the single cell, there is a hope to build a model of cellular networks that will reflect actual tissue properties.

Here and after we refer spontaneous EEG measured on the scalp as EEG and μV as measure units if they are not specified. Normal EEG waveforms are defined and described by the frequency, amplitude, and location [54], [121]. The most common clinical classification [56] for sleep EEG based on frequency characteristic and presented in Table 2.1 and Figure 2.3.

Table 2.1: Basic EEG activity.

Activity	**Frequency range**
Alpha	8–13 Hz
Beta	$\geq$13 Hz
Delta	0–3.99 Hz
Theta	4–7.99 Hz
Sigma	11–16 (most commonly 12–14) Hz

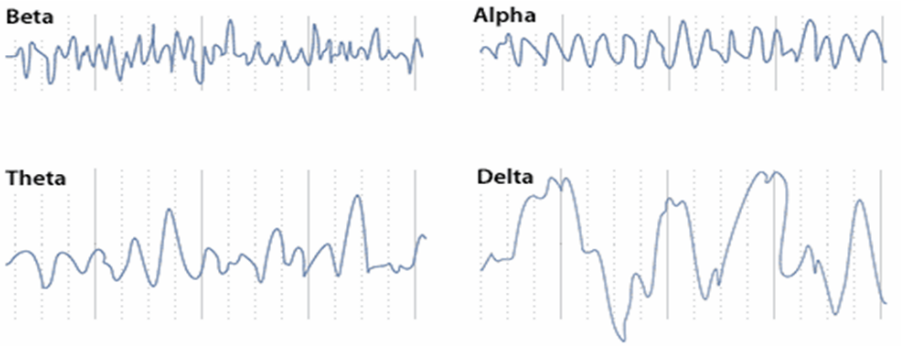


Figure 2.3: Examples of basic EEG waves from [79].

2.1.2 Polysomnogram

Polysomnogram (PSG) is a comprehensive recording of the biophysiological processes that occur during sleep. PSG is usually performed at night during sleep. Such recordings can last several hours. This diagnostic test monitors many body functions. The American Academy of Sleep Medicine (AASM) [56] recommendations propose to obtain the following activity:

- Left and right mastoid processes at channels M1 and M2 respectively. Channel M1 is usually used as a reference electrode for sleep scoring in recorded EEG. If M1 fails during the recording, backup electrodes should be used and referenced to M2.
- Brain activity, EEG. At least F4, C4, O2 (or/and backup F3, C3, O1) channels should be recorded. EEG is obtained by referencing to M1. In clinical practice

there are often recorded all available channels (e.g. 19 channel by system 10-20 depicted in Figure 2.1).

- Eye activity, electrooculogram (EOG). Two electrodes, E1 and E2 are measured. E1 is placed 1 cm below and 1 cm lateral to the left outer canthus, and E2 is placed 1 cm above and 1 cm lateral to the right outer canthus. These electrodes pick up the activity of the eyes in virtue of the electropotential difference between the cornea and the retina. EOG is obtained by referencing E1 and E2 to M2 (M1 in case M2 fails).

- Muscle activity, electromyogram (EMG). It is recommended to record EMG using three electrodes placed at the chin. Ch_z is placed in the midline 1 cm above the inferior edge of the mandible. And two electrodes are placed 2 cm below the inferior edge of the mandible. Ch_2 is 2 cm to the right, and Ch_1 is 2 cm to the left. Final EMG is obtained as Ch_1-Ch_z and Ch_2-Ch_z.

The recommended sampling frequency is 500 Hz and 200 Hz as a minimum. Additionally to the listed channels there may be recorded electrocardiogram (ECG), body temperature, airflow, oximetry, nasal pressure, limb EMG, and many other body measurements, which may describe a patients state during sleep.

2.2 Human sleep

Sleep is the resting state of the body. Its main function is organism restoration [18]. Sleep quality directly affects physical [85] and mental health [14] as well as cognitive performance [34], [86]. Daily cycles of sleep and wakefulness are controlled by circadian rhythm [117]. An analysis of body measurements during sleep became the main investigation tool in sleep medicine because sleep is an unconscious process (though, subjective sleep evaluation is still very useful for sleep research and clinical diagnosis). Objective sleep evaluation in sleep clinics is based on sleep PSG analysis. A recording device is applied to an investigated subject before sleep and removed when sleep is over, and the subject is awake again. Recorded data are scored into non-overlapped time windows and are marked as one of four categories called sleep stages or wake based on activity presented in PSG and especially EEG. Further, the biological data are studied and explained by underlying physiological processes.

2.2.1 Sleep stages

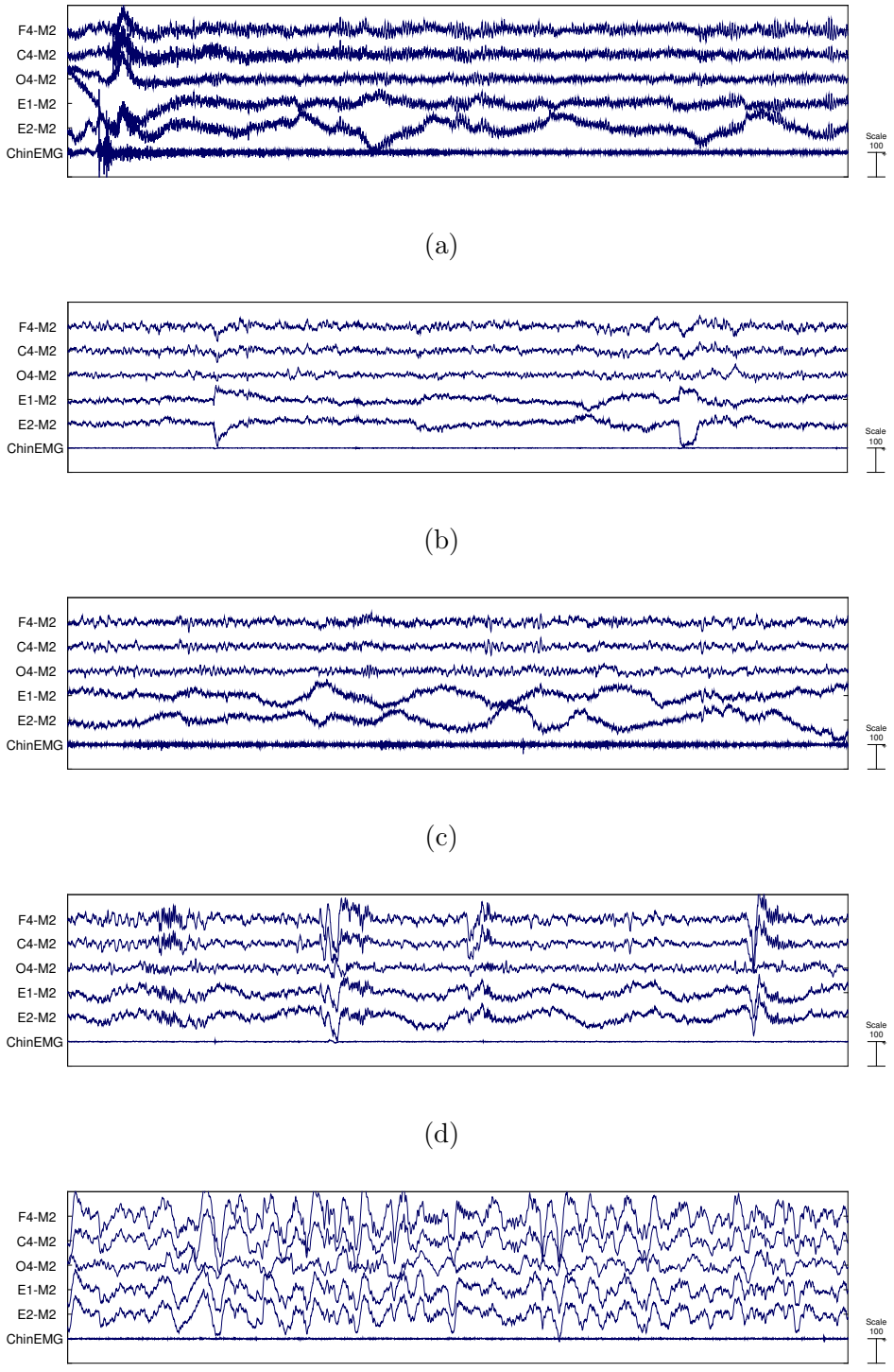
Sleep stages are evaluated in short-time epochs. AASM sleep scoring manual [56] states length of a scoring window as 30 s and previously used Rechtschaffen & Kales guidelines [102] recommended 20 s window. Hallmarking sleep stage characteristics in EEG and PSG within an epoch differ for infants, children, and adults. Here and after we refer to scoring rules for adults by AASM scoring manual. Sleep stages and wake activity characteristics are listed below:

- Wakefulness. Alpha rhythm is prevailing in EEG during wake state in the occipital region. Eye movements are recorded in EOG and produce eye blinking and eye movement artifacts in EEG. The chin EMG during this stage is of variable amplitude but is usually higher than during sleep. Example is provided in Figure 2.4a.

- Rapid eye movement (REM) sleep stage. REM sleep accounts for 20–25% of sleep time. It is also referred to as paradoxical sleep because EEG during the REM sleep is similar to that of waking in EEG. However, in comparison with wakefulness, it has low chin EMG tone. The first REM sleep episode occurs 60–90 min after the sleep onset. EEG tracings during REM sleep are characterized by a low-voltage, mixed-frequency activity with slow alpha (defined as 1–2 Hz slower than wake alpha) and theta waves. Example is provided in Figure 2.4b.

- Non-rapid eye movement (non-REM) 1 sleep stage. This comprises 3–8% of sleep time. Stage 1 sleep occurs most frequently in the transition from wakefulness to the other sleep stages or following arousals during sleep. In stage 1 non-REM sleep, alpha activity (8–13 Hz), which is characteristic for wakefulness, diminishes and a low-voltage, a mixed-frequency pattern emerges. The highest amplitude electroencephalography activity is generally in the theta range (4–8 Hz). EMG activity decreases, and EOG demonstrates slow rolling eye movements. Vertex sharp waves may be present toward the end of stage 1 non-REM sleep. Example is provided in Figure 2.4c.

- Non-REM 2 sleep stage. It begins after approximately 10–12 min of stage 1 non-REM sleep for a healthy subject and comprises 45–55% of total sleep time. The characteristic EEG features of stage 2 non-REM sleep include sleep spindles and K-complexes. A sleep spindle is described as a short burst in a sigma frequency band. A K-complex is a waveform with two components: a negative wave followed by a positive wave, both lasting not more than 0.5 s. The EMG activity is diminished in comparison to wakefulness. Example is provided in Figure 2.4d.



(a)

(b)

(c)

(d)

(e)

Figure 2.4: 30-s epochs of PSG for one subject during wakefulness (**a**), REM (**b**), non-REM 1 (**c**), non-REM 2 (**d**) and non-REM 3 (**e**) sleep stages.

- Non-REM 3 sleep stage. This stage corresponds to a slow wave sleep because of the presence of moderate amounts of high-amplitude slow-wave activity. Muscle tone is decreased in comparison to wakefulness or stage 1 non-REM sleep. Sleep spindles

may occur in this stage. It occupies 15–20% of total sleep time. The previous guidelines [102] divided this stage into non-REM 3 and 4 stages by the amount of slow wave activity. Example is provided in Figure 2.4e.

Period of sleep, which corresponds to stages non-REM 1 and 2 is called light sleep. Non-REM 3 stage is also known as deep sleep, which is connected to regeneration processes [18], [85].

2.2.2 Sleep architecture

Sleep architecture can be represented by a graph called hypnogram. An example is shown in Figure 2.5. It represents how sleep stages as well as arousal and wakefulness follow each other during the night. Further, we explain a typical sleep architecture for a healthy subject with normal sleep. It consists of four to five sleep cycles, sometimes called episodes. A sleep cycle begins with a short period of non-REM 1 stage progressing through non-REM 2, followed by stage non-REM 3, back to non-REM 2 and finally to REM. The average length of the first cycle is 70–100 min. The second, and later, cycles are longer, approximately 90–120 min [65]. Proportions of sleep stages differ across the night. In normal adults, REM sleep increases as the night progresses and is longest in the last one-third of the sleep cycle. As the sleep cycle progresses, non-REM 2 begins to occupy the majority of non-REM sleep, and non-REM 3 may disappear [25]. In healthy adults sleep architecture may be affected by age [75], [103], gender [75] and pregnancy and the postpartum period in women [104].

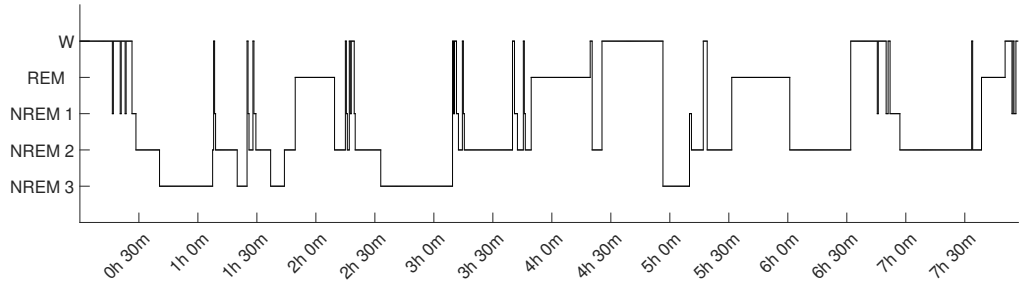


Figure 2.5: Hypnogram example. Shifting between wake state (W) and non-REM 1 (NREM 1), 2 (NREM 2), 3 (NREM 3) and REM (REM) sleep stages for one healthy subject over one night.

2.2.3 Sleep disorders

The International Classification of Sleep Disorders third edition (ICSD3) by American Academy of Sleep Medicine [118] presents basic division of over 100 different sleep

diagnoses into the following categories:

- Insomnia. It is one of the most common sleep disorders. It reveals itself in difficulty falling asleep or staying awake, even when a person has the chance to do so. It could be caused by mental issues like stress or abnormality in a brain structure. ICSD3 divides patients suffering from insomnia into three groups: chronic, short-term, and others. EEG recordings of insomniacs show a significant increase in the beta power frequency range in non-REM and REM [83], which could be evidence of excessive hyperarousals during sleep. Also, specific prefrontal sleep pattern during the whole sleep period may be observed in such patients [98].

- Sleep-related breathing disorders. They are divided into four sections: obstructive sleep apnea, central sleep apnea syndrome, sleep-related hypoventilation disorder, and sleep-related hypoxemia disorder by ICSD3. It is diagnosed by criterion frequencies of breathing disturbance using measurements from an oronasal thermal airflow sensor to monitor airflow [56]. Some disorders of this type affect sleep architecture in terms of longer sleep latency [11].

- Central disorders of hypersomnolence. They are caused by intrinsic central neural system abnormalities in control of sleep-wake generators. Narcolepsy belongs to this group. In recorded EEG non-REM sleep stages are mixing with wakefulness. Cataplexia and sleep paralysis are often events observed in sleep recordings. Simultaneous occurrence of REM sleep-related muscular paralysis and sudden, brief muscular weakness [77]. It is detected by analysis of muscle activity in the chin EMG [56].

- Circadian rhythm sleep-wake disorders. The main syndrome of such disorder is the inability to sleep in the desired time. It may be connected with the dysfunction of the inner biological clock or affected by environmental and social reasons (shift work, jet-lag). Sleep recording shows no abnormalities [77].

- Parasomnias. This type includes diagnosis of unwanted behavior during sleep. The common feature characterizing this group is dissociated sleep stages where sleep stages mix. It includes both EEG and non-EEG features of the sleep stage. One example is sleepwalking, where wakefulness mix into non-REM sleep. Another example REM sleep behavior disorder (RBD) shows a mixture of wakefulness and REM in recorded data. One of its hallmarks is increased muscular tone during REM [39], [99]. Clinicians differ phasic and tonic EMG activity. First one is represented as short burst lasting 0.1–5 s and more than twice as high as the background

amplitude. Tonic activity is characterized as a constant increase by at least a factor of two or four compared to baseline amplitude.

- Sleep-related movement disorders. Restless legs syndrome and periodic limb movement disorder fall into that category. Periodic limb movements (PLMs) are scored using electrodes placed on limbs by technical specification described in [56]. PLMs are scored as a series of at least four 0.5–10 s bursts in EMG limb electrodes.

2.3 Sleep EEG patterns

Detection of two kinds of EEG patterns is discussed in the study. First one is an artifact pattern, which can be defined only by comparison to normal EEG patterns. The other typical kind of EEG pattern is described by spectral and amplitude characteristics. We concentrate on sleep spindle detection. In the following subsections, we describe their characteristics, provide visual examples, and discuss previously proposed methods of automatic detection of these EEG patterns.

2.3.1 Artifact detection

Events not related to brain activity are called artifacts. Typical causes of artifacts are technical issues (detached electrodes, sweating, electrical noise) and physiological events (muscle contractions, movements, eye rolling) of the body. For instance, artifacts caused by blinking are characterized by slow-frequency/high-amplitude waves that can be seen mainly in frontal electrodes. The typical waveform corresponding to such artifacts is shown in Figure 2.6a. Blinking occurs only during the awake state. Artifacts associated with movements may arise during both being awake and asleep, the latter divided into REM, non-REM 1, 2, and 3 sleep stages [56]. Such artifacts are characterized by unusually high amplitude and unusual frequency in several electrodes, as it is shown in Figure 2.6b. Electrode artifacts may arise during the whole recording, and they may have different characteristics. Voltage jumps could be seen in all channels and are illustrated in Figure 2.6c. A loss of contact of electrodes may produce noise in a channel or high amplitude anomalies, as shown in Figure 2.6d.

The main reason for artifact detection is that they may noticeably affect further processing, make it impossible or lead to false results. Manual data inspection for artifacts is a demanding and tedious task, especially in long-term recordings such as a whole-night EEG. Moreover, the number of electrodes and their locations are typically set according to the areas of interest. Modern sleep research often utilizes at least 19 channels distributed evenly over the head. Some recent studies have applied up to 256-channel high-density

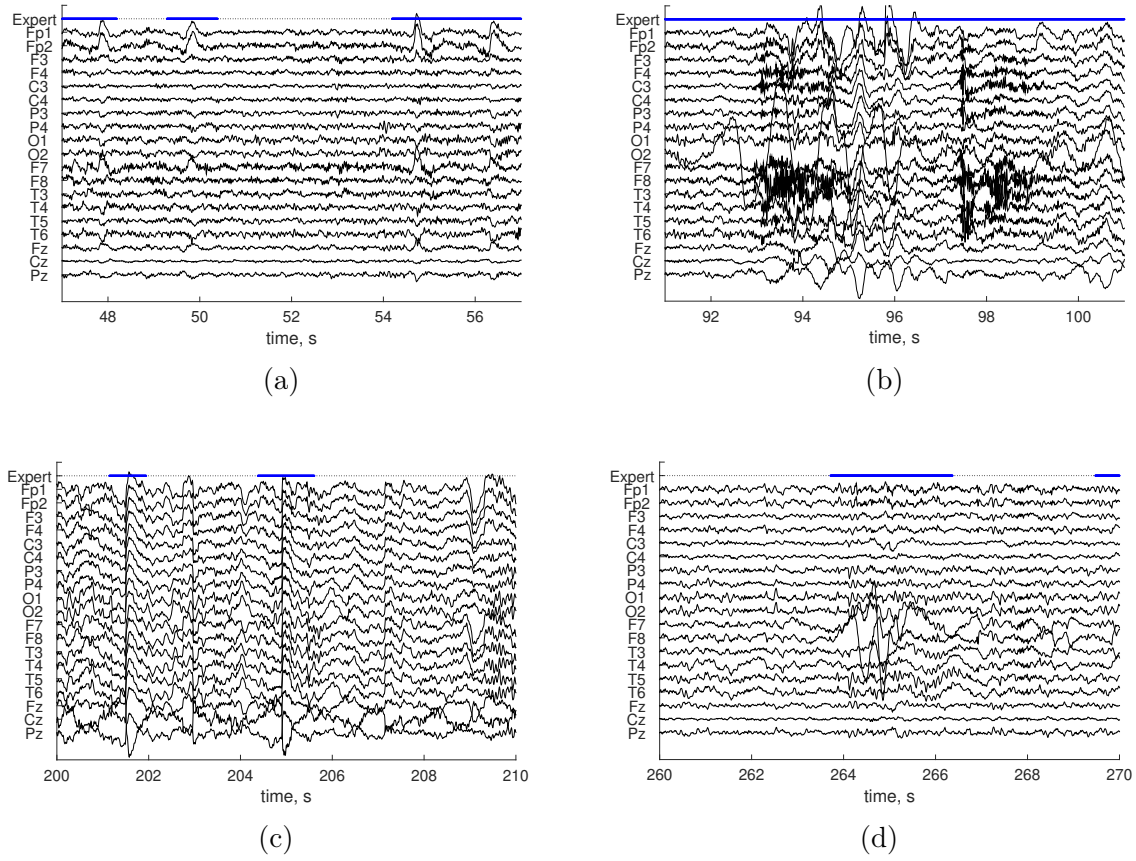


Figure 2.6: Expert labeling (Expert) of blinking artifacts (**a**), movement artifact during wakefulness (**b**), and electrode artifacts in non-REM 3 (**c**) and REM (**d**). Solid blue lines are the expert's artifact marks.

EEG [100], [120]. That makes manual artifact detection a time-consuming task requiring specific skills. There is a wide choice of automated artifact detection methods available in the literature [59]. They include approaches based on signal decomposition such as Fourier transform [122], wavelet transform [8], [123], or empirical mode decomposition [15], [23]. Methods based on regression [17], [63] are popular for eye blinking artifact detection. Blind source separation (BSS) methodologies like independent component analysis (ICA) [59], [82], [122] and independent vector analysis [22] are often used for ocular, muscular, and cardiac artifact rejection. Once the decomposition has been achieved, components corresponding to artifact activity may be rejected manually or using automatic methods [19]. A combination of BSS and other signal decomposition techniques may bring benefits for muscular artifact removal [21], [24]. Several recent studies have proposed methods based on classification techniques such as random forest classifiers [8], artificial neural networks [92], or support vector machine classifiers [33], [70], where training examples determine the types of artifacts to be detected. Some studies focus on optimization of existing approaches [2], [47].

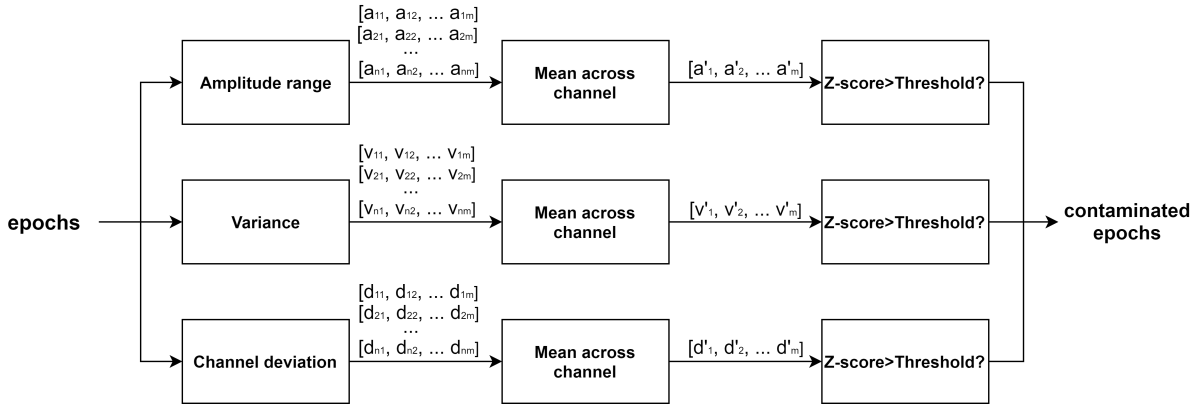


Figure 2.7: FASTER detection of single-epoch artifacts.

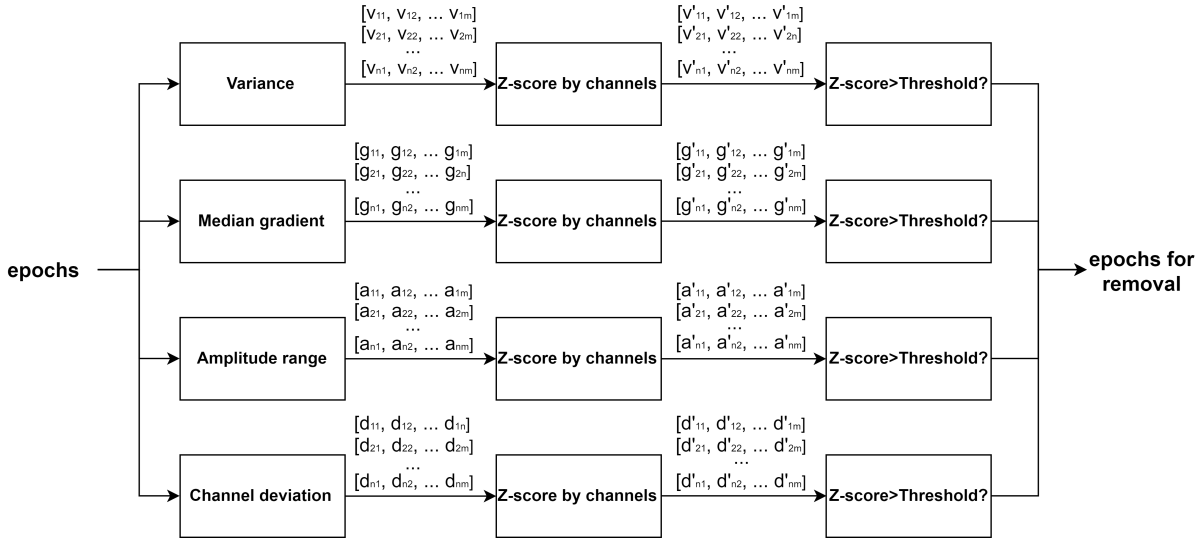


Figure 2.8: FASTER detection of single-channel, single-epoch artifacts.

Among the artifact detection strategies for wake EEG processing, we can emphasize a fully automated, unsupervised method for processing of high density EEG data (FASTER). It was cited over 250 times in total (including 68 times in 2018 and over 25 times in 2019). It covers maximum artifacts, relatively simple and adaptive. Not less important is that the authors provide a free open-source plugin for EEG Lab software [30]. The method works with epochs of equal size. It has many steps, including ICA decomposition, channel, and epoch rejection. However, here were concentrate on two main steps for the epoch rejection. One of them rejects epochs based on comparison with other epochs, and it is schematically represented in Figure 2.7. Another step is based on comparison channels within one epoch. It is depicted in Figure 2.8. They are based on the same principle: (i) extract characteristic features, (ii) apply z-score transformation defined in Section 2.4.2 and (iii) outlier detection by a thresholding of z-score values. The authors originally set the z-score threshold at 3. More details on extracted features are provided in Section 2.4.2.

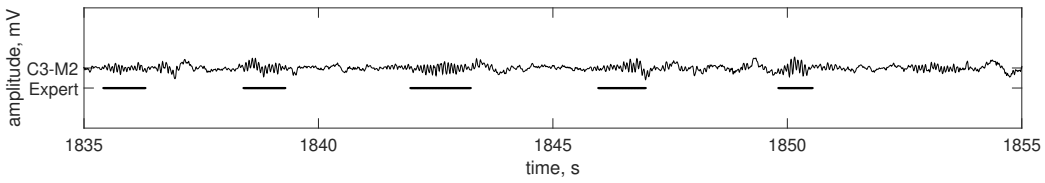


Figure 2.9: Example of sleep spindles in C3-M2 EEG channel (C3-M2) and expert labeling (Expert).

Research in the field of sleep data processing mainly focuses on the detection and elimination of the most expected artifacts during sleep like muscular and pop-out artifacts [128] through single-channel approaches [78]. These methods concentrate only on certain artifact types. Therefore, artifact-free data are obtained after sequential application of different methods, and single-channel approaches must be applied to all channels alternately.

2.3.2 Sleep spindle detection

A sleep spindle is a pattern in sleep EEG characterized as a 0.5–2-s long burst in 11–16 Hz frequency range. An example of sleep spindles scored by an expert is provided in Figure 2.9. Some studies differentiate fast and slow spindles by the prevailing frequency and their location [45]. In sleep medicine, such patterns hallmark non-REM 2 sleep stage [56]. However, they may also occur during slow-wave sleep [45]. Also, sleep spindles are connected to circadian modulation [64]. Crucial sleep spindle parameters such as number, duration, amplitude, frequency, and density are investigated in the study of cognitive abilities in children [20] and adults [105] like learning [43] and memory consolidation [16]. However, sleep spindle parameters are highly dependent on a subject. Some factors like age [45], [80], and some mental disorders [41], [130] are associated with changes in sleep spindle development and, consequently, their parameters. Sleep spindle density is an important index in studies of brain and psychological disorders like schizophrenia [40], [129], epilepsy [89], and autism [72].

Manual detection of sleep spindles is a challenging task, even for experts. Analysis of hours of EEG data makes it tedious in the first place. Moreover, despite the strict definition, sleep spindles are hard to distinguish in sleep EEG. Therefore, agreement of different clinical experts may differ for the same data [95]. However, agreement of trained experts favorably differs from non-experts agreement [31], [73], [131]. All these reasons lead to the development of automatic methods of sleep spindle detection. The most general approach for the detectors is to apply a detection function sensitive to sleep spindle and then define them by application of a threshold [5], [31], [55], [80], [88]. Some of them will be described further. The threshold may be stated before or calculated from the recorded EEG ac-

tivity. Often, the detection function utilizes the information present in a corresponding frequency band. Accessing to frequency band corresponding to sleep spindles is performed using band-pass filtering [80], [88], S-transform [55], empirical mode decomposition [110], [132] and wavelet transform [132]. Several methods for classification were proposed, such as artificial networks [127] and support vector machine [67]. However, the requirement of prepared examples dataset makes it difficult to use in real practice. Modern methods are mostly focused on improving existing methods using advanced approaches [60], [68] including optimization [73], [96], [116] and combination of methods [37]. Several popular methods for automatic sleep spindle detection are depicted in Figure 2.10.

In the thesis, we concentrate on four sleep spindle detectors. Each one of them is based on a different detection function. The first sleep spindle detector based on root mean square (RMS) utilizes root mean square (described in Section 2.4.1) of amplitude of the filtered signal [88]. The second sleep spindle detector using a relative sigma power (RSP) was proposed in [31]. The third sleep spindle detector utilizing S-transform and defining sigma index computed for obtained frequencies (SIGMA) was proposed and tested on a private database [55]. The last one is a sleep spindle detector based on Teager energy operator (TEO) [5], which is described in Section 2.4.1. Detection of sleep spindles is performed by application of a threshold. There were used implementations of the methods by [95] and [131] and threshold values proposed in [95].

2.4 Automatic sleep pattern detection

There are two main approaches to the automatic EEG pattern localization. In the first approach, a signal is transformed, and detected events are identified using a threshold which could be previously stated or obtained from the recorded data. Methods using threshold are widespread [31], [40], [87], [93] and relatively easy to implement. However, the problem of the threshold may arise with changes in recording procedure or subject group specificity. For example, sleep spindle detection in young and adult subjects require different sensitivity of the threshold used by method [95].

The other approach utilizes machine learning techniques. Classification methods utilizing support vector machine (SVM), Naive Bayes classifier, neural networks or combinations of methods [1], [3], [74], [110] were published. These techniques use data with EEG patterns scored by an expert for model training. In medicine, collecting such data examples is problematic because manual EEG pattern labelling is not a part of typical clinical routine and it is a time-consuming task requiring specific skills. Moreover, analysis of data collected from subjects with specific features (age, gender, sleep disorder, etc.) should be performed with a model trained on data with the same characteristics.

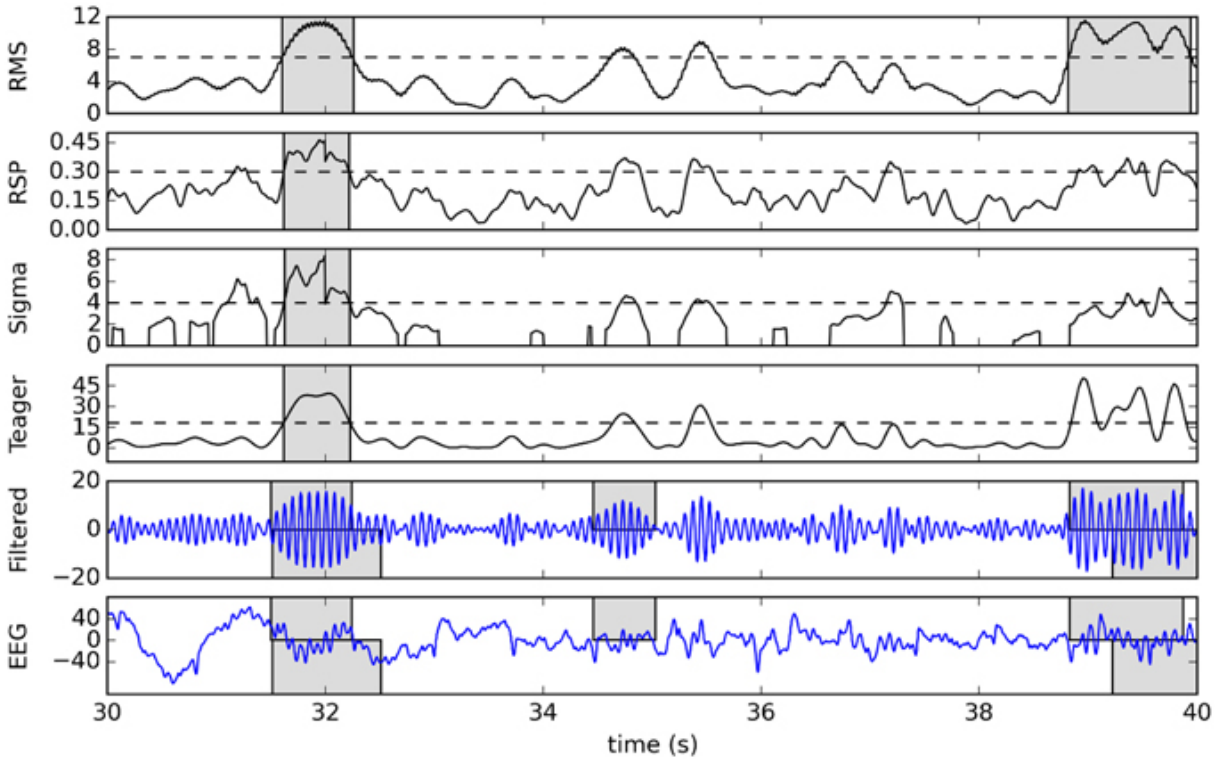


Figure 2.10: Example of detection using a 10-s sample from open Dreams database from [95]. The original signal and its 11–16 Hz band-passed version are plotted in the two bottom graphs, with gray boxes showing expert scoring (top rectangles for V1, bottom for V2). The four plots in the upper portion of the figure show corresponding to TEO (Teager), SIGMA (Sigma), RSP (RSP) and RMS (RMS) detection functions (solid lines), effective thresholds (dashed lines), and detected spindles (gray boxes).

However, there exists another approach to data pattern localization. An unsupervised approach, clustering, could be employed. It assumes that all data epochs represented in a feature space could be grouped into separate clusters. This approach is unsupervised and estimates similarity between feature vectors to group the data into clusters. Patterns are detected using cluster analysis under assumptions stated by the domain knowledge or previously published research. This approach assigns data to clusters separately for every data recording which allows the model to adapt to the data better and overcome variability between subjects.

The method overview is presented in Figure 2.11. In the first step, the raw signal is filtered and divided into smaller segments or epochs. Filtration and segmentation methods are presented in Section 2.4.1. Feature extraction transforms the segments to the feature space, that is described in Section 2.4.2. Detection is performed using clustering methods explained in Section 2.4.3. Detected events are further analyzed by experts. Often, it includes visual inspection and further processing and analysis. It may be computing of important properties in case of clinically significant EEG patterns or events rejection in case of artifacts. The method could be evaluated using scoring provided by an expert.

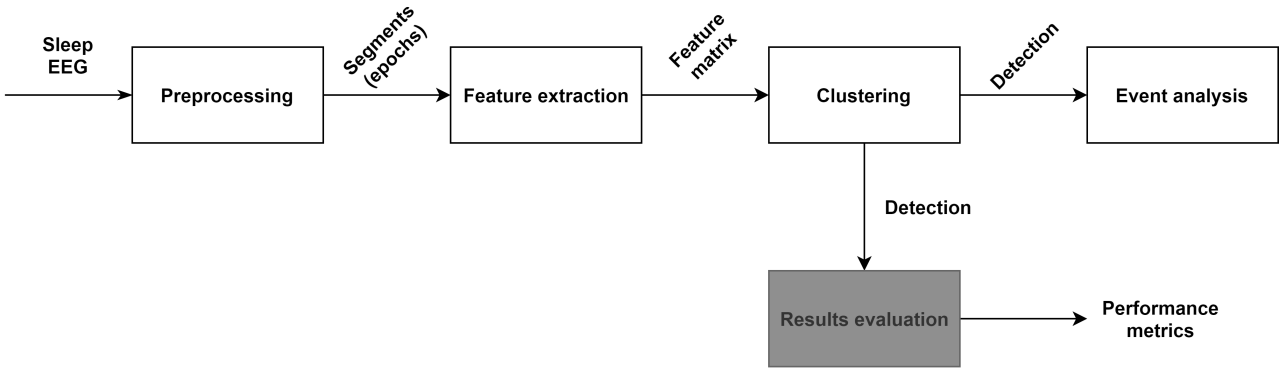


Figure 2.11: Automatic EEG pattern detection method overview.

Detailed explanation of performance evaluation is provided in Section 2.4.4.

2.4.1 Preprocessing

Filtration and transform

The goal of the EEG filtration is to eliminate irrelevant activity based on the frequency criteria. First power-line noise elimination should be mentioned. It affects 50 Hz in Europe and 60 Hz in the United States. The notch filter is used for elimination. This filter rejects a narrow frequency band and leaves the rest of the spectrum almost undisturbed [57].

Moreover, filtration is used to remove frequencies not relevant to the investigated ones. Such as sleep relevant frequencies are considered in 0.3–35 Hz [56]. However, upper border of this range may vary in 30–35 Hz range. High pass filters are used for low-frequency removal such as baseline shift; low pass filters remove high-frequency noise. Moreover, such filtering could be used for investigation of narrower bands. For example, a sigma band could be used for sleep spindle detection and a gamma band for arousal analysis. In this study, we do not concentrate on filtration problem and use a standard filtration method provided by the EEG Lab software. It uses a finite impulse response filter with a coefficient of three times sampling frequency divided by lower cut off frequency by default.

Transformation of a filtered signal could emphasize the signal characteristics. For reinforcement of sudden changes in the filtered signal, a Teager energy operator Ψ [6], [66] could be applied to the signal. This signal transformation method is beneficial for signal segmentation. It estimates the instantaneous energy of the signal. It is defined for a discrete signal (x_n is a signal sample) as

$$\Psi[x_n] = x_n^2 - x_{n-1}x_{n+1}. \tag{2.1}$$

For the signal smoothing we use convolution $f(t) = s * h(t)$, where t is a time, f is a smoothed signal, s is an original signal and h is a smoothing kernel. Hamming window of

suitable length normalized by a sum of coefficients could be used as a smoothing kernel.

Segmentation

The task of signal segmentation consists of splitting a signal into smaller epochs (segments). Smaller epochs tend to have more consistent properties, which makes the processing more effective. Wide range of methods formally requires this condition, including fast Fourier transform or ICA. Segmentation is particularly important and frequently used for long-term EEG analysis because of its long duration and extremely nonstationary properties.

There are two main approaches to signal segmentation: constant and adaptive. Constant segmentation is described by two parameters: window size and overlap. It divides the signal into pieces with the same length. Obtained segments are often called epochs. In contrast, adaptive segmentation adjusts the segment size to the signal change points. It cuts the signal only when signal statistic changing is "noticeable enough". The signal statistic is usually calculating in a sliding window. The most popular algorithm is named after its author Alpo Värri [125] and it is based on computing frequency and amplitude measures in two successive sliding windows. A more sophisticated algorithm uses a fractal dimension of the signal in a sliding window as a feature for segmentation [10], [38]. Other methods utilize non-linear energy operator [42], [61]. These methods have linear complexity, can work online, and are very suitable for long-term signal analysis. Other approaches are based on signal prediction. If a mismatch between the prediction and the original signal is higher than a defined threshold, then it suggests a potential segment boundary.

Let $[x_1, x_2, x_3, \ldots, x_{N-1}, x_N]$ be the signal of length N and let $w_i = [x_i, x_{i+1}, x_{i+2}, \ldots, x_{i+L-1}]$ be a window of length L starting at each sample of the signal. A function M evaluates a signal statistic in a window w_i. Further, one window is evaluated by M every K steps so that one obtains $J + 1$ values

$$[M(w_1), M(w_{1+K}), M(w_{1+2K}), \ldots, M(w_{1+JK})].$$

The segment border detection is based on evaluation of the absolute difference between two successive windows j and $j + 1$:

$$G_j = |M(w_{1+jK}) - M(w_{1+(j-1)K})|, \tag{2.2}$$

where $j \in \{1, \ldots, J\}$. Thus, a sequence of J values for all $J + 1$ windows is obtained:

$$G = [G_1, G_2, ..G_J], \tag{2.3}$$

which reflects the change of statistical properties of the signal. G is further normalized by division of each G_j by $\max_j G_j$. The border detection is performed by thresholding the G, detecting local maxima, and simultaneously satisfying two other constraints. The distance between the peaks must be higher than the window size L and the amplitude of the peak must be higher than the standard deviation of the thresholded signal. The adaptive threshold is used, which is obtained as a moving average value of the G sequence in window L^* multiplied by threshold coefficient c.

In comparison to constant segmentation, the adaptive approach has much more parameters: the window length L, the step of the evaluation function K, threshold coefficient c and threshold window L^*. The window length parameter L is essential. It should be large enough to detect the difference in the two windows at all, but it should not be too large, which could avoid identifying some borders by capturing a lot of real segments in one window. The proper window length can be chosen using the energy of the G sequence. Ideally, G values with the appropriate window size should be highly above zero close to segment boundaries and almost zero elsewhere. This property can be evaluated using energy value [10]. For an improper window length, G function has more energy compared to a proper window length. Thus, for an analyzing window with length L, the energy of the G sequence can be calculated as its mean squared value [10]:

$$E = \frac{\sum_{j=1}^{J} {G_j}^2}{J}, \tag{2.4}$$

where J is the length of the G sequence. A proper window length should correspond to the minimum of the energy curve.

2.4.2 Feature extraction

After EEG signal preprocessing, the information needed for classification (features) is extracted from the segments. A well-chosen feature should remain unchanged if variations take place within a class, and it should reveal significant differences when discriminating between patterns of different classes. Examples of features often used to describe EEG [46], [52] include statistical features (mean, standard deviation, kurtosis, skewness, median, Hjorth parameters), features computed for typical EEG bands (absolute or relative spectra), entropy-based features (Shannon or spectral entropy), features connected with the signal similarity and self-similarity (correlation, auto-correlation or covariance).

Further, we specify some features relevant to the study. We denote $x = [x_{s+1}, \ldots, x_{s+k}]$ as a segment of signal $X = [x_1, \ldots, x_N] \in R^N$, it starts at s and has duration of k samples, $s + k < N$. Then we can define

- root mean square value as $\sqrt{\frac{1}{k}\sum_{i=1}^{k} {x_{s+i}}^2}$;

- amplitude range as $\max(x) - \min(x)$;
- mean value μ as $\frac{1}{k}\sum_{i=1}^{k} x_{s+i}$;
- standard deviation (std) σ as $\sqrt{\frac{1}{k}\sum_{i=1}^{k}(x_{s+i}-\mu)^2}$, where μ is mean value of x defined above;
- variance as σ^2;
- median slope as median value of the first derivative of the signal;
- channel deviation (from channel mean) as $(\frac{1}{N}\sum_{i=1}^{N} x_i) - \mu$, where μ is mean value of x defined above [93];

For a signal segment $y = [y_{t+1}, \ldots, y_{t+k}]$ of signal $Y \in R^M$, which starts at t and has duration of k samples, $t + k < M$. We can define sample Pearson correlation coefficient r as

$$r = \frac{\sum_{i=1}^{k}(x_{s+i}-\mu_x)(y_{t+i}-\mu_y)}{\sqrt{\sum_{i=1}^{k}(x_{s+i}-\mu_x)^2}\sqrt{\sum_{i=1}^{k}(y_{t+i}-\mu_y)^2}}, \tag{2.5}$$

where μ_x and μ_y are mean values of x and y respectively. Values p range from –1 to +1, a value 0 indicates no correlation.

Sometimes features should be normalized in order to avoid possible problems caused by insufficiently scaled features. One of the most popular ways of normalization is z-scoring. Each recalculated value x'_i is obtained from the original x_i by subtracting a mean value $\bar{x}$ of the feature and division by standard deviation σ_x of the feature.

$$x'_i = \frac{x_i - \bar{x}}{\sigma_x}. \tag{2.6}$$

2.4.3 Clustering

The aim of the clustering method is to divide a set of N observations into k clusters in such way that elements within one cluster are more similar to each other than to elements in other clusters. Clustering analyses a set of observed feature vectors $X = \{x_1, \ldots, x_N\}$ where $x_i \in R^M$, M is a number of extracted features. Then, the clustering outcome a vector $L = [l_1, \ldots, l_N]$ where each $l_i \in \{1, \ldots, k\}$ shows cluster assignment for vector x_i.

K-means

K-means [4] is a straightforward clustering method. It divides the data into k clusters where k is a predefined parameter. The idea of the method lies in iterative refinement.

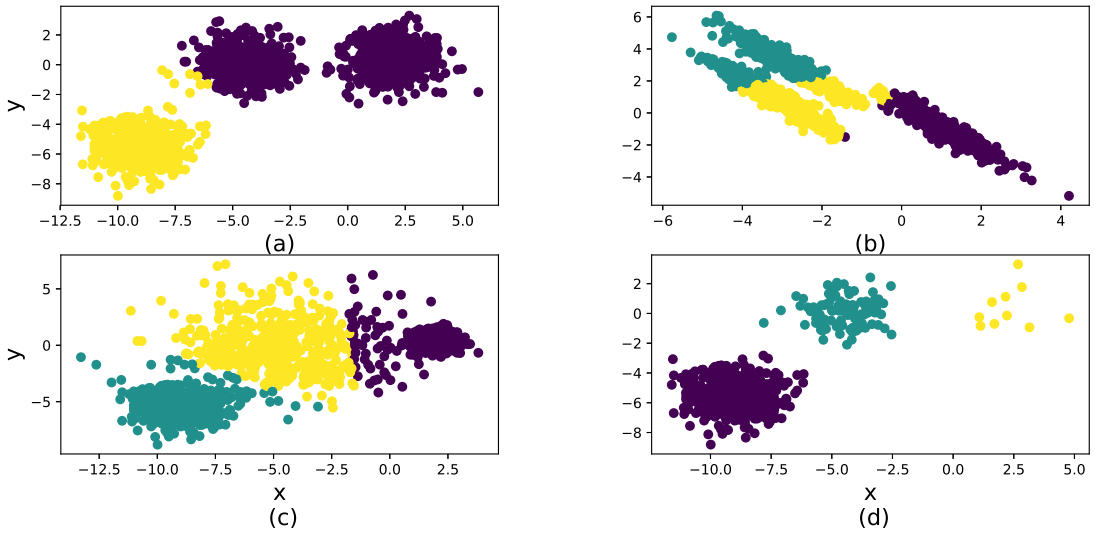


Figure 2.12: K-means outcome in case of incorrect number of clusters (**a**), anisotropically distributed clusters (**b**), unequal variance (**c**) and unevenly sized clusters (**d**). Elements are represented by features x and y. Different colors represents different clusters.

It solves the optimization task of minimizing the sum of squared within-cluster distances. Each iteration consists of two steps. First, observations are assigned to a cluster by the closest centroid. Then, new centroids are updated based on a new cluster partition. The distance between elements can be defined by any distance function. Centroids are usually calculated as cluster means based on distance function definition. For the most common variation of the method, Euclidean distance and mean values as cluster centroids are utilized. The method converges to a local minimum, therefore, starting initialization matters and multiple runs are preferred to find the best solution.

The biggest advantage of k-means is simplicity. It is also suitable for large datasets in terms of computational time and clustering efficiency. However, the biggest disadvantage is the problem of the optimal k. Inappropriate k value leads to inefficient clustering as it is shown in Figure 2.12a. Moreover, k-means is very sensitive to data. First of all, it is very sensitive to scale. An example of anisotropically distributed data is in Figure 2.12b. In the standard implementation, it assumes that data are distributed into hyper-spherical clusters with the same size. Therefore, it has benefits on data, which fit this assumption. Example of clusters with unequal variance are in Figure 2.12c. Nevertheless, it is stable to unevenly sized clusters when data assumption is fulfilled as it is shown in Figure 2.12d.

Gaussian mixture model and expectation maximization algorithm

Gaussian mixture model (GMM) [4] assumes that the observed data come from k independent Gaussian distributions. We denote the parameters of the distribution in cluster

j as $\Theta_j = \{\mu_j, \Sigma_j\}$. The aim of GMM is to identify parameters Θ_j and, consequently, determine clusters. We introduce a latent variable z_i for x_i. It randomly chooses from $\{1, \ldots, k\}$ and shows, from which cluster x_i comes. Variable z_i is drawn from multinominal distribution where $p(z_i = j) = \alpha_j$, $\alpha_j \geq 0$ and $\sum_{j=1}^{k} \alpha_j = 1$. We specify a joint distribution $p(x_i, z_i) = p(x_i|z_i)p(z_i)$.

The parameters of our model $\Theta = \{\alpha_1, \ldots, \alpha_k, \mu_1, \ldots, \mu_k, \Sigma_1, \ldots, \Sigma_k\}$ are estimated using log likelihood of the data:

$$\ell(\Theta) = \log p(X; \Theta) = \sum_{i=i}^{N} \log\ p(x_i; \Theta) = \sum_{i=1}^{N} \log \sum_{j=1}^{k} p(x_i, z_j; \Theta). \tag{2.7}$$

assuming that x_i are independent observations. Estimation of $\ell(\Theta)$ explicitly might be difficult. Expectation maximization (EM) iterative algorithm is an effective way to find maximum likelihood estimates of parameters of our probabilistic model. It consists of two steps. In the E-step, it constructs a lower bound on ℓ, and then optimizes that lower-bound in the M-step.

For each i, let the Q_i be some distribution over z's, $\sum_z Q_i(z) = 1$, $Q_i(z) \geq 0$.

$$\sum_i \log\ p(x_i; \Theta) = \sum_i \log \sum_{z_i} p(x_i, z_i; \Theta) = \sum_i \log \sum_{z_i} Q_i(z_i) \frac{p(x_i, z_i; \Theta)}{Q_i(z_i)}. \tag{2.8}$$

Considering that $f = \log x$ is a concave function, $f'' < 0$ and that

$$\sum_{z_i} Q_i(z_i) \Big[\frac{p(x_i, z_i; \Theta)}{Q_i(z_i)} \Big]$$

is an expectation of $[p(x_i, z_i; \Theta)/Q_i(z_i)]$ with respect to z_i drawn according to distribution given by Q_i [91]. Then, we can derive from Jensen's inequality

$$\sum_i \log p(x_i; \Theta) = \sum_i \log \sum_{z_i} Q_i(z_i) \frac{p(x_i, z_i; \Theta)}{Q_i(z_i)} \geq \sum_i \sum_{z_i} Q_i(z_i) \log \frac{p(x_i, z_i; \Theta)}{Q_i(z_i)}. \tag{2.9}$$

In this way, we successfully construct the lower bound for $\ell(\Theta)$. Using $Q_i(z_j) = p(z_j|x_i; \Theta)$ allows us to squeeze the lower bound to obtain the equality [91]. Then, the parameters Θ should be chosen to maximize the following quantity

$$\sum_i \sum_{z_i} Q_i(z_i) \log \frac{p(x_i, z_i; \Theta)}{Q_i(z_i)} \tag{2.10}$$

Then, EM iteration for GMM model and a dataset X in M-dimensional space will be as following

E-step: obtain "guesses" of membership of x_i from X in a mixture component j as $w_{ij} = Q_i(z_i = j) = p(z_i = j|x_i; \Theta)$. Using Bayes rule we get

$$w_{ij} = p(z_i = j|x_i; \Theta) = \frac{p(x_i|z_i = j; \Theta)\ p(z_i = j; \Theta)}{\sum_{c=1}^{k} p(x_i|z_i = c; \Theta)\ p(z_i = c; \Theta)}. \tag{2.11}$$

Obtained w_{ij} provides the soft assignments to the clusters.

M-step: update parameters values of

$$\begin{aligned}\sum_i \sum_{z_i} Q_i(z_i) \log \frac{p(x_i, z_i; \Theta)}{Q_i(z_i)} &= \sum_{i=1}^{N} \sum_{j=1}^{k} Q_i(z_i = j) \log \frac{p(x_i|z_i = j; \Theta)p(z_i = j; \Theta)}{Q_i(z_i = j)} \\ &= \sum_{i=1}^{N} \sum_{j=1}^{k} w_{ij} \log \frac{\frac{1}{\sqrt{(2\pi)^M |\Sigma_j|}} \exp(-\frac{1}{2}(x_i - \mu_i)^T \Sigma_j^{-1}(x_i - \mu_j)\alpha_j}{w_{ij}}\end{aligned} \tag{2.12}$$

by taking derivative with respect to μ_j, Σ_j and, finally, α_j and setting it to zero. The final obtained rules are [91]

$$\alpha_j = \frac{1}{N} \sum_{i=1}^{N} w_{ij}, \tag{2.13}$$

$$\mu_j = \frac{\sum_{i=1}^{N} w_{ij} x_i}{\sum_{i=1}^{N} w_{ij}}, \tag{2.14}$$

$$\Sigma_j = \frac{\sum_{i=1}^{N} w_{ij}(x_i - \mu_j)(x_i - \mu_j)^T}{\sum_{i=1}^{N} w_{ij}}. \tag{2.15}$$

Initialization starts with random parameters. The Jensen's inequality and choice of Q_i's ensure that EM always monotonically improves log-likelihood [91]. A common practice to monitor the convergence is to test if the difference of log likelihoods at two successive iterations is less than a predefined tolerance parameter.

The problem of optimal k is still open for this clustering method. Example is depicted in Figure 2.13a. The GMM is more stable to specific structures of data as shown in Figure 2.13b and Figure 2.13c. Another problem, which may arise, is that insufficient number of observation in a mixture component leads to inaccurate estimation of distribution parameters. Figure 2.13d provides an example.

DBSCAN

Density-based spatial clustering of applications with noise (DBSCAN) [4] is another clustering method. The idea is in grouping nearby elements into a cluster and mark elements in low-density regions as outliers. It requires two parameters ϵ and τ. By analyzing

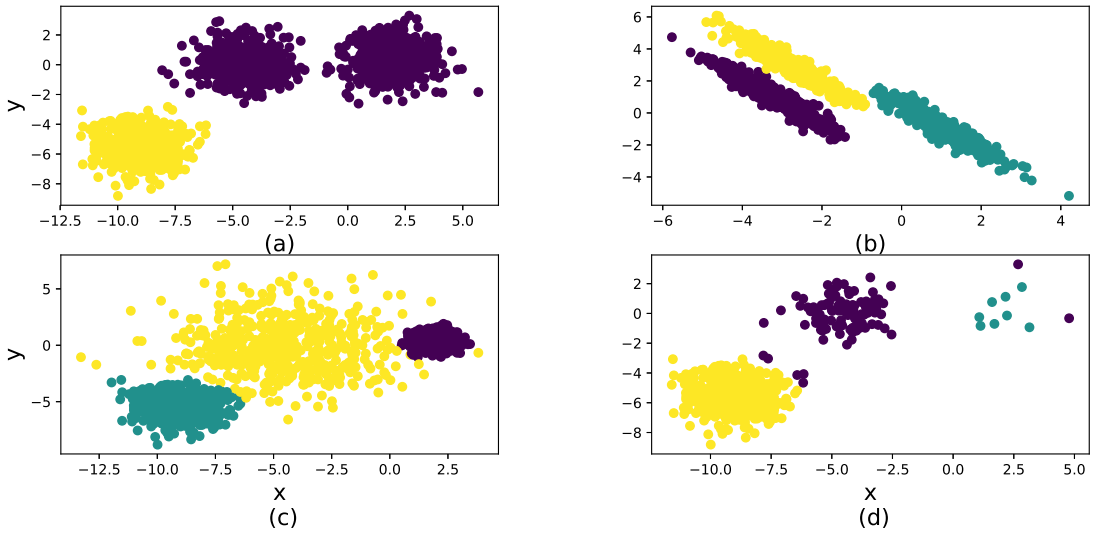


Figure 2.13: GMM outcome in case of incorrect number of clusters (**a**), anisotropically distributed clusters (**b**), unequal variance (**c**) and unevenly sized clusters (**d**). Elements are represented by features x and y. Different colors represents different clusters.

ϵ-neighborhood of each point we assign it to be (i) a core point, which neighborhood contains τ points, (ii) a border point, which neighborhood contains less than τ points but at least one core point, and (iii) an outlier, which neighborhood contains less than τ points and no core points. Then, connectivity graph is constructed with respect to the core points. A node corresponds to a core point and edge is added between a pair of core points only when distance between them is less than ϵ. All connected components correspond to clusters. Border points are assigned to clusters with which they have the highest connectivity level [4]. Parameters are set by the domain expert. However, if data are not well understood, setting these parameters might be difficult. In comparison to clustering techniques described above, the number of clusters is not required, which gives advantages in applications. Example is shown in Figure 2.14a. Moreover, cluster forms is not specified by the clustering method and robust to outliers as it is depicted in Figure 2.14b and Figure 2.14d. Nevertheless, large differences in densities may cause ineffective clustering outcome (Figure 2.14c).

2.4.4 Results validation

A typical procedure for validation of pattern detection results is based on a comparison of two binary vectors of the same size. One is obtained using automatic detection; the second one is provided by an expert. Specific performance measures are calculated to assess the efficiency of the detection. They are all based on four basic values true positive (TP), true negative (TN), false positive (FP), and false negative (FN), whose estimation

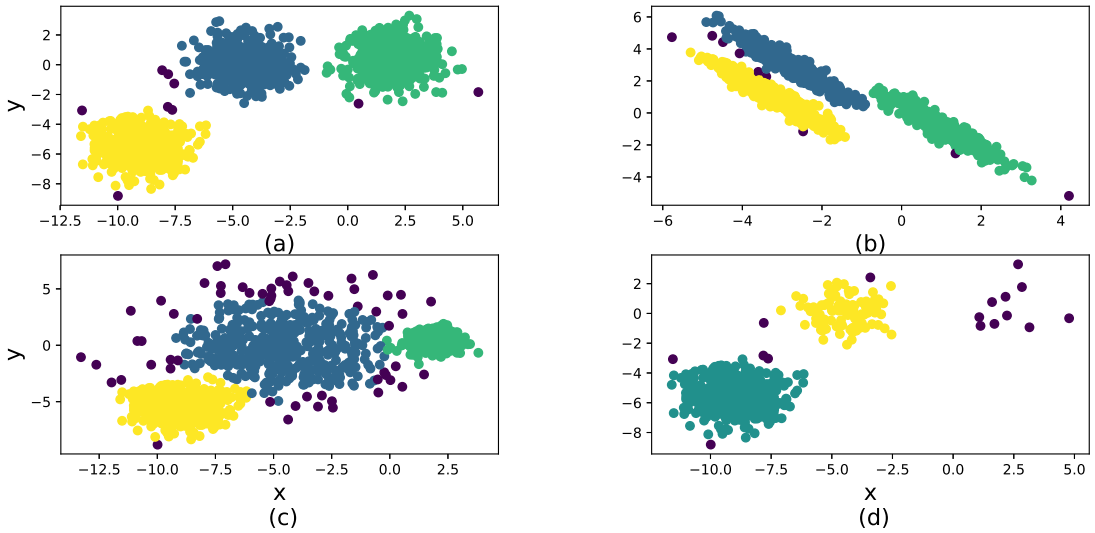


Figure 2.14: DBSCAN ($\epsilon = 0.3, \tau = 10$) outcome in case of three clusters (**a**), anisotropically distributed clusters (**b**), unequal variance (**c**) and unevenly sized clusters (**d**). Elements are represented by features x and y. Dark purple indicates outliers and other colors represents different clusters.

is dependent on the task. Typical examples are true positive rate (TPR) ($\frac{\mathrm{TP}}{\mathrm{TP+FN}}$) and true negative rate (TNR) ($\frac{\mathrm{TN}}{\mathrm{TN+FP}}$). Further, we describe two main strategies used in the study.

In the case of unsupervised learning, validation is performed on each available data signal separately. Obtained results may be grouped by data origin or important data properties. Often, information about how an event of interest is detected is more important than the ability of a method to separate data signals in two classes. Therefore, we introduce two strategies of result evaluation. For simplicity let call them *Strategy 1* and *Strategy 2*. *Strategy 1* concentrates on evaluation of event discovery efficiency. Then, validation of results is performed by comparison to labeling provided by the expert. We denote

- true positives TP as the number of sleep spindle events labeled by an expert, which were marked as sleep spindles by a detector at least for a 0.3 s;
- false positives FP as the number of detected events not counted in TP;
- false negatives FN as the number of sleep spindle events labeled by an expert and not counted in TP;

We use recall *Rec* and precision *Prec* and their harmonic mean $F1$ to measure effec-

tiveness of the proposed method as

$$Rec = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{2.16}$$

$$Prec = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{2.17}$$

$$F1 = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}. \tag{2.18}$$

Metrics are obtained on an event basis. That makes the evaluation strategy more appropriate for the detection of characteristic patterns like sleep spindles.

In some cases, we might be interested in measuring ability of a method to separate data signals in two classes like in case of artifact detection. In *Strategy 2*, we measure goodness of detection in both positive and negative classes. Similarly to a previous case, performance evaluation is obtained by comparing the automatic detection to a scoring provided by human experts. Here we describe the case of epoch-based evaluation. For comparison of continuous detections, evaluation is performed on a sample-by-sample basis. We denote the number of true positives, false positives, true negatives, and false negatives as TP, FP, TN, and FN, respectively.

Following [128], we compute Cohen's kappa K. It is considered a robust measure because it takes into account agreement occurrence by chance. Values of K greater than 0.6 are considered good, and values higher than 0.8 indicate an excellent agreement [69]. We have:

$$K = \frac{P_o - P_r}{1 - P_r}, \tag{2.19}$$

where:

$$P_o = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}, \tag{2.20}$$

$$P_r = \frac{(\mathrm{TP} + \mathrm{FN})(\mathrm{TP} + \mathrm{FP}) + (\mathrm{FP} + \mathrm{TN})(\mathrm{TN} + \mathrm{FP})}{(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})^2}. \tag{2.21}$$

Sensitivity Se and false discovery rate FDR were also compared in this strategy. We use different notations TPR, Se and Rec to distinguish evaluation strategies.

$$Se = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{2.22}$$

$$FDR = \frac{\mathrm{FP}}{\mathrm{TP} + \mathrm{FP}}. \tag{2.23}$$

Chapter 3

Artifact Detection in a Sleep EEG Processing Pipeline

3.1 Introduction

This chapter discusses the problem of artifact detection in a long-term multichannel sleep EEG as a part of a data processing pipeline in a real sleep research center. The described pipeline is used by sleep investigators after the data are processed at the clinic, and sleep stages are detected. Therefore we assume that the sleep stages are revealed. Artifact detection is performed before further investigation such as spectral analysis or other artifacts sensitive processing. Since EEG data are very complex, the final decision must be controlled manually by an expert after automatic detection. The goal of automatic processing is to detect as many artifacts as possible to relieve investigators work.

In this chapter, we employ several previously proposed artifact detection techniques as artifact detection engine of the pipeline. We focus on recently introduced methods such as automatic artifact detection in full-night PSG [128], and selected methods that demonstrated good performance for rejection of 20-s epochs of single-channel sleep EEG [78]. Additionally, we test a popular state-of-the-art method for wake EEG, FASTER [93]. Moreover, we propose an adaptive improvement for FASTER. In comparison to FASTER utilizing the definition of an outlier and calculating z-scored statistics, we use DBSCAN for clustering of FASTER based features to define outliers. We assume that the artifact-free data group forms a high-density cluster. DBSCAN algorithm solves the clustering problem, and the biggest detected cluster is assigned to artifact-free data. This approach allows us to adapt to the data of each subject separately and overcome the problem of inter-subject variability. We adapt all previously mentioned methodologies for requirements of the pipeline and testing data. We test the method on full-night EEG recordings from subjects suffering from insomnia. The data contain a wide variety of presented

artifacts. Cohen's kappa coefficient is used to access the performance of detection.

The method presented in this chapter is a part of the sleep PSG processing pipeline used by the sleep research group at the National Institute of Mental Health, Czech Republic. The pipeline was used for the preparation of publication submitted for a review to a journal with impact factor: D. Dudysova, K. Veldova, M. Smotek, E. Saifutdinova, J. Koprivova, J. Buskova, B. A. Mander, M. Brunovsky, P. Zach, J. Korcak, V. Andrashko, M. Viktorinova, F. Tyls, A. Bravermanova, T. Froese, T. Palenicek, and J. Horacek, "Effects of daytime administration of psilocybin on sleep: Similar changes to antidepressants?", *Frontiers in Pharmacology*, 2019.

3.2 Materials and methods

3.2.1 Data

The data used in the chapter were obtained from the National Institute of Mental Health, Czech Republic. The whole-night EEG as a part of the PSG was recorded from subjects suffering from insomnia. Total number of recordings is 13. Each recording was obtained using 19 channels placed according to the 10-20 system: Pz, Cz, Fz, T6, T5, T4, T3, F8, F7, O2, O1, P4, P3, C4, C3, F4, F3, Fp2, Fp1. Channels were referenced to an average of EEG channels by the choice of experts. The sampling frequency was 250 Hz. Sleep stages were scored by a trained clinician. Data were filtered in frequency range 0.5–40 Hz on the preprocessing step using EEG lab filtering procedure.

The data were divided into 5-s epochs. Epochs containing artifacts were labeled by the experts by visual inspection of EEG performed separately for each recording in the EEG Lab software using function "reject continuous data by eye" [30]. The expert was instructed to label all visible artifacts. Evaluation excludes the whole wake stage since it was not relevant for the study, for which data processing was performed. Detection of eye movement artifacts in REM was not precise since ICA was planned for ocular artifact rejection. That explains unexpectedly small artifact rate in REM. Details on the dataset is provided in Table 3.1.

Table 3.1: Dataset details.

Sleep Stage	Length (epochs)	Artifact Rate (%)
all sleep stages	3699.2 ± 1117.1	0.12 ± 0.08
REM	750.5 ± 302.3	0.15 ± 0.13
non-REM 1	263.5 ± 121.6	0.21 ± 0.09
non-REM 2	1818 ± 588	0.13 ± 0.09
non-REM 3	924.9 ± 270.9	0.07 ± 0.08

3.2.2 Sleep EEG processing pipeline

The overview of the processing pipeline for a long-term sleep EEG is presented in Figure 3.1. A similar pipeline is used in several world sleep research laboratories. The 19-channel EEG as a part of PSG is recorded. Channels are referenced to an average of all EEG channels. Sleep scoring is performed by trained clinicians. Processing assumes the previous filtering is important for sleep frequency range and removal of defective channels. First, the continuous EEG data are divided into 5-s non-overlapped epochs. Automatic artifact rejection method labels the epochs using provided sleep scoring. Due to the high complexity of the data, manual evaluation by visual inspection is performed after the automatic processing and the outcome represents a final decision before further analysis. The artifact-free epochs are used for further processing. Further processing could be performed in epochs such as spectral analysis or continuous signals such as sleep spindle detection.

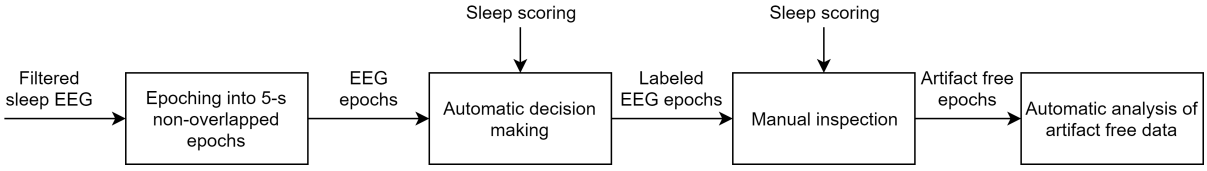


Figure 3.1: Sleep EEG processing pipeline.

Artifact detection in a long-term EEG is a very laboratory specific task. Many of the methods presented in the literature [78], [128] process raw data and steps of analysis includes proper filtering, channel rejection, and contaminated epoch detection. Often they assume that sleep stages are not known. However, here data were previously inspected by clinicians. Obtained data are filtered, sleep stages, and a list of channels to reject are provided. Therefore, we concentrate on the detection of contaminated epochs of length 5 s. We can adapt well-known strategies like FASTER, use rules for contaminated epoch as in [128] or in [78], or apply a proposed adaptive strategy based on FASTER features explained in Section 3.2.3. In the next section, we adopt all previously proposed

methodologies for requirements of the pipeline and testing data.

The pipeline presented in this chapter is developed on the basis of EEG Lab software version not older than 13 for MATLAB. Programming code for manual inspection step is provided below. EEG structure from EEG Lab contains data in a 3-dimensional matrix $D \in R^{M \times K \times N}$ where M is a number of channels, K is a length of epoch in samples and N is a number of epochs. Redefining *reject.rejmanual* data structure field with rejection binary vector *reject* of length N we save it into the data structure. Rejection vector is obtained by any automatic method. It requires redefining of *reject.rejmanualE* field with binary matrix E of size $M \times N$ where each value E_{ij} denotes whether channel i at epoch j is rejected (1 is rejection and 0 is not). Working with EEG Lab structure saves the time on calculation of exact borders of labeled events. In our pipeline researchers were interested in epoch rejection not channels, therefore, we defined E as a zero matrix. We plot EEG using function *pop_eegplot*. Each window contains 18 epochs, which is 1.5 min by default. Channel signals are plotted with 100 μV spacing. We set a new reject-button command using variable *command*. Reject-button function saves a new manually corrected rejection vector into *reject.mat* file. In the next steps this file is automatically loaded and used in analysis.

```
[nch, nsam, nep] =  size(EEG.data);
EEG.reject.rejmanual = reject;
EEG.reject.rejmanualE = zeros(nch, nep);

command = ['[nch, nsam, nep] =  size(EEG.data);'...
'[trialrej elecrej] = eegplot2trial( TMPREJ, nsam, nep, []);' ...
'reject = trialrej; save(''reject.mat'', ''reject'');'];

pop_eegplot(EEG, 1, 0, 0, '', 'spacing', 100, 'winlength', 18, 'command'
, command);
```

3.2.3 Proposed method: AFAST

A proposed adaptive artifact detection method based on FASTER features (AFAST) is presented in this section. In the first stage, we remove non-EEG epochs corresponding to severe electrode artifacts. They are characterized by extremely high correlation between EEG channels. Therefore, for every epoch, we obtain an average correlation between channels. Correlation is obtained as a Pearson correlation coefficient. Epochs with a value over 0.9 are removed. This trick allows us to reduce the number of contaminated epochs in further processing. Figure 3.2 illustrates example of the data and corresponding average correlation presented in Figure 3.3.

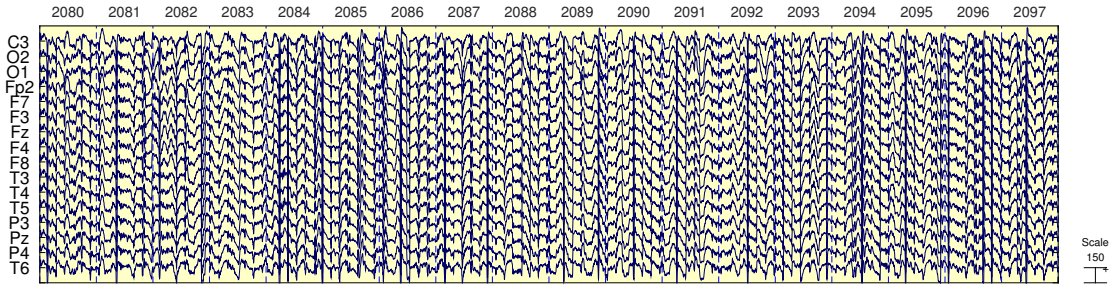


Figure 3.2: Recorded EEG with severe electrode artifacts. Vertical dashed lines separate 5-s epochs, epoch numbers are given above. Contaminated epochs scored by an expert are highlighted with yellow color.

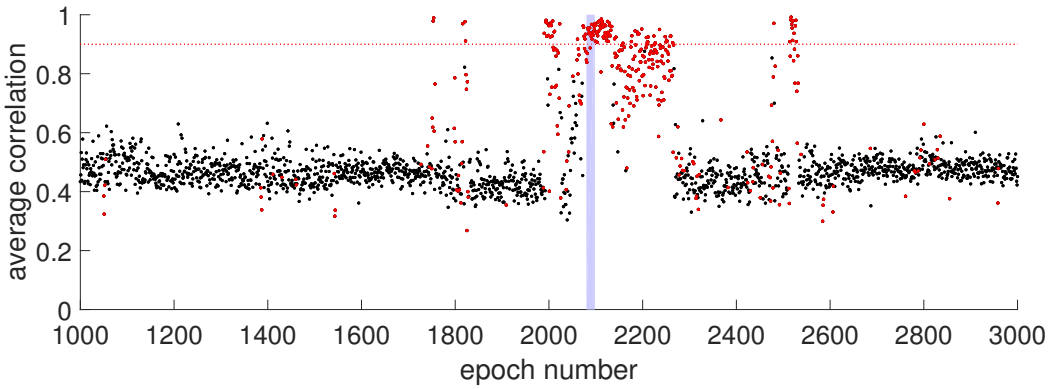


Figure 3.3: Average correlation between channels in a epoch is represented by a black dot for artifact-free 5-s epoch and by a red dot contaminated 5-s epoch. Blue area corresponds to epochs shown in Figure 3.2.

In the next step we extract features from each epoch. We use features proposed in FASTER for epoch rejection. Exactly, from each i_{th} epoch $e_i \in R^{M \times K}$ where K is an epoch length and M is a number of channels we extract:

- channel amplitude range in each epoch channel $a_i \in R^M$, we reduce a_i to features $am_i = max(a_i)$ and $ad_i = max(a_i)/median(a_i)$;
- channel variance in each epoch channel $v_i \in R^M$, we reduce v_i to features $vm_i = max(v_i)$ and $vd_i = max(v_i)/median(v_i)$;
- channel deviation from channel mean in each epoch channel $d_i \in R^M$, we reduce d_i to features $dm_i = max(d_i)$ and $dd_i = max(d_i)/median(d_i)$.

Feature details are provided in Section 2.4.2. Features are normalized by median value across the set. We represent epoch e_i in three R^2 spaces by (am_i, ad_i), (vm_i, vd_i) and (dm_i, dd_i). Plotting that spaces reveals that artifact-free epochs form a high-density area

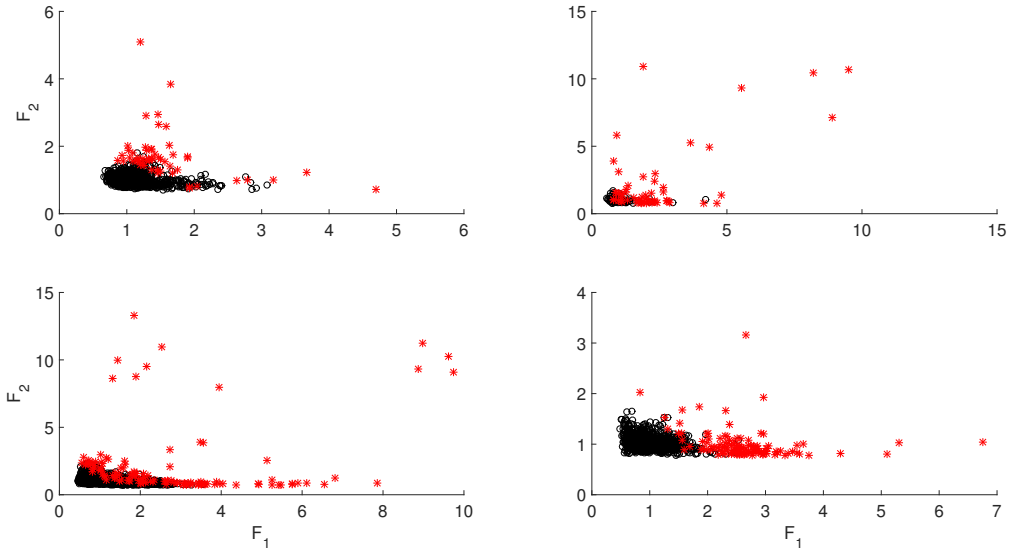


Figure 3.4: Features distribution for different sleep stages for one subject. Each i_th epoch is represented by am_i ($Feature_1$) and ad_i ($Feature_2$). Artifact-free and contaminated epochs are denoted by black and red color respectively.

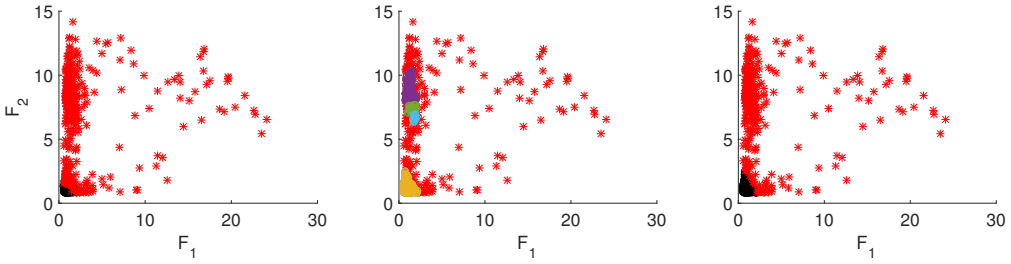


Figure 3.5: Complex distribution of features. Each i_th epoch is represented by am_i ($Feature_1$) and ad_i ($Feature_2$). Expert scoring (**a**), outcome of DBSCAN (**b**) and AFAST (**c**). In (a) and (b) artifact-free and contaminated epochs are denoted by black and red color respectively. In (c) light blue, dark blue and yellow represents different clusters and red are outliers.

whereas contaminated epochs lay separately from it as shown in Figure 3.4. Moreover, artifact-free distribution is not normal due to the presence of different EEG patterns. It is especially noticeable in non-REM 2 stage where there are both high-amplitude K-complexes and low amplitude activity due to long 5-s duration epochs containing the majority of K-complexes and other overlaps in a feature space. Application of clustering methods, which assumes cluster distribution like GMM, does not bring many benefits. DBSCAN detects clusters by density distribution of features. Therefore, it is able to detect clusters of artifact data as well as shown in Figure 3.5.

Before the clustering procedure, we reduce the feature set on all epochs i such as $am_i < median(am_i)$, to avoid false positive results. We perform DBSCAN clustering in each feature space for sleep stage separately. Distance between instances is determined as

Euclidean distance. We set 10 as a minimum number of instances to set a core point and average distance to 10_{th} element over all elements as a radius of investigated neighborhood. A cluster of artifact-free data is determined by the maximum number of elements. In case of several such clusters, we choose one with the smallest average value across am_i. All elements, which do not belong to an artifact-free cluster, we label as artifacts.

3.3 Experiment results

In this section, we compare automatic epoch rejection methods. Testing was performed on dataset described in Section 3.2.1. All contaminated electrodes were removed before testing. Performance evaluation utilizes epoch based *Strategy 2* from Section 2.4.4 to discover ability of a method to separate data signals in two classes. Cohen's kappa K, sensitivity Se, and false discovery ratio FDR were calculated for the whole recording and different sleep stages separately. Additionally, we obtained the percentage of epochs labeled as artifacts.

First, artifact detection strategy proposed for fMRI artefact rejection and sleep scoring toolbox (FASST) [128] was applied for contaminated epoch detection. The slope threshold was reduced to 1.5×10^3 $\mu V/s$ to reduce a big number of FN. Detection of muscular artifacts was performed in 1-s epochs and then extended to 5-s scoring by the following rule. If in 5-s epoch there is at least one 1-s epoch labeled as an artifact, then the whole 5-s epoch is labeled as an artifact. Obtained results are presented in Table 3.2. Next, we adopted artifact detection strategy based on single channel artifact detection methods (SCADM) described in [78]. We chose all methods with adaptive threshold and a method which show the best performance. Selected methods were simultaneously applied for artifact detection. Two methods based on thresholding of maximal amplitude and slope were selected into SCADM. An adaptive threshold value is calculated as an std of the parameter multiplied by a predefined coefficient. We used 6.2 for amplitude and 4.3 for slope thresholding instead of the values proposed in the paper to reduce the number of FN. Such values were determined as optimal using grid search for the whole dataset. Band power threshold was also included in single channel methods as a method exhibiting the best performance in the original publication. Each 5-s epoch with band power at any channel greater than 8.4 μV/Hz is labeled as an artifact. Results are presented in Table 3.3. Then, we utilized FASTER for detection of 5-s contaminated epochs. Z-scored values are obtained with stable statistics like median and median absolute deviation. The threshold applied to z-scored values is 6 for non-REM 3 and 4.5 for other sleep stages instead of originally used 3 to reduce FP. Results are shown in Table 3.4. Finally, we used AFAST for detection of contaminated epochs. Results are presented in Table 3.5.

K values obtained for all sleep stages are visually compared in Figure 3.6. Note that analysis by all four methods was performed in sleep stages separately.

Table 3.2: Performance of FASST artifact detection.

Sleep Stage	$\boldsymbol{K}$	$\boldsymbol{Se}$	$\boldsymbol{FDR}$	Artifact Rate (%)
all sleep stages	0.32 ± 0.15	0.24 ± 0.14	0.22 ± 0.16	0.04 ± 0.05
REM	0.37 ± 0.14	0.30 ± 0.14	0.21 ± 0.15	0.05 ± 0.07
non-REM 1	0.27 ± 0.16	0.23 ± 0.13	0.28 ± 0.22	0.06 ± 0.04
non-REM 2	0.31 ± 0.20	0.23 ± 0.18	0.18 ± 0.18	0.04 ± 0.06
non-REM 3	0.27 ± 0.23	0.22 ± 0.21	0.34 ± 0.29	0.02 ± 0.05

Table 3.3: Performance of SCADM artifact detection.

Sleep Stage	$\boldsymbol{K}$	$\boldsymbol{Se}$	$\boldsymbol{FDR}$	Artifact Rate (%)
all sleep stages	0.44 ± 0.17	0.37 ± 0.15	0.30 ± 0.22	0.07 ± 0.07
REM	0.45 ± 0.22	0.36 ± 0.23	0.14 ± 0.19	0.08 ± 0.12
non-REM 1	0.44 ± 0.19	0.44 ± 0.17	0.30 ± 0.24	0.14 ± 0.07
non-REM 2	0.44 ± 0.21	0.37 ± 0.21	0.26 ± 0.20	0.07 ± 0.07
non-REM 3	0.29 ± 0.25	0.33 ± 0.27	0.52 ± 0.38	0.05 ± 0.06

Table 3.4: Performance of FASTER artifact detection.

Sleep Stage	$\boldsymbol{K}$	$\boldsymbol{Se}$	$\boldsymbol{FDR}$	Artifact Rate (%)
all sleep stages	0.47 ± 0.11	0.40 ± 0.14	0.21 ± 0.11	0.05 ± 0.02
REM	0.54 ± 0.21	0.46 ± 0.20	0.11 ± 0.11	0.05 ± 0.02
non-REM 1	0.41 ± 0.17	0.37 ± 0.13	0.22 ± 0.20	0.09 ± 0.02
non-REM 2	0.47 ± 0.11	0.44 ± 0.16	0.26 ± 0.19	0.06 ± 0.02
non-REM 3	0.45 ± 0.28	0.38 ± 0.27	0.17 ± 0.12	0.02 ± 0.02

Table 3.5: Performance of AFAST artifact detection.

Sleep Stage	$\boldsymbol{K}$	$\boldsymbol{Se}$	$\boldsymbol{FDR}$	Artifact Rate (%)
all sleep stages	0.70 ± 0.10	0.80 ± 0.11	0.29 ± 0.18	0.13 ± 0.7
REM	0.67 ± 0.24	0.83 ± 0.21	0.27 ± 0.29	0.17 ± 0.13
non-REM 1	0.73 ± 0.12	0.74 ± 0.17	0.14 ± 0.10	0.19 ± 0.11
non-REM 2	0.72 ± 0.16	0.89 ± 0.08	0.30 ± 0.23	0.15 ± 0.08
non-REM 3	0.61 ± 0.19	0.66 ± 0.27	0.29 ± 0.23	0.06 ± 0.7

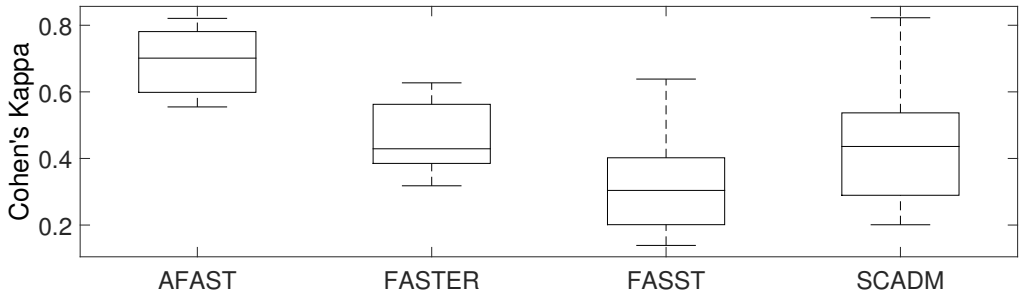


Figure 3.6: Distribution of Cohen's kappa for different artifact detection methods.

Performance of the first employed strategy FASST is much lower in comparison to the other methods. FASST was designed for detection of the pop-out electrode and movement artifacts. Visual inspection of data reveals that channel loss of contact artifacts (example is depicted in Figure 2.6d) are very common for the recordings. However, FASST is unable to detect them. Moreover, it labels a very small amount of epochs; the median value of detected epochs across recordings is 67. That makes detection very sensible for false detections. The majority of FP are epochs with the beta increase but not labeled by an expert. The number of such epochs is 16.6 ± 12.2. Such proportions noticeably increase FDR and, consequently, sufficiently affect sensitivity and, therefore, K values.

The next employed strategy SCADM shows better performance in terms of Se and K. It detects artifacts by abnormal slope, amplitude, and frequency. The main impact in performance, however, comes from thresholding by band power, which was established as a constant threshold. However, in the case where such a threshold is too low, it produces plenty of false positive results. It can be observed in REM and non-REM 1 especially. Besides, the method makes false detections by an amplitude criteria in non-REM 2 and 3 by detecting normal high amplitude delta activity. The discussed method generates a great variability in the results. SCADM achieves the best performance in recordings full of pop-up electrodes. Instead of using a constant value as it is in FASST, it adapts to the data and adjusts the threshold using std value. Such an approach allows the method to achieve better results overall.

AFAST and FASTER method are based on the same features. Results obtained for AFAST show significant improvement against all other methods in terms of K and Se. FASTER provides stable results over the dataset but shows much lower performance. Because of too many artifacts in one channel, it becomes less sensitive to artifact detection due to statistics used in z-scoring, and only extreme values are labeled as artifacts. An example is provided in Figure 3.7. There are too many artifacts occurring in channel Pz, and only extreme values are labeled as artifacts. Otherwise, if there are not so many artifacts in the channel, it tends to generate too many FP. Due to this reason and a small number of epochs, FASTER labels as artifact each epoch with amplitude increase in non-

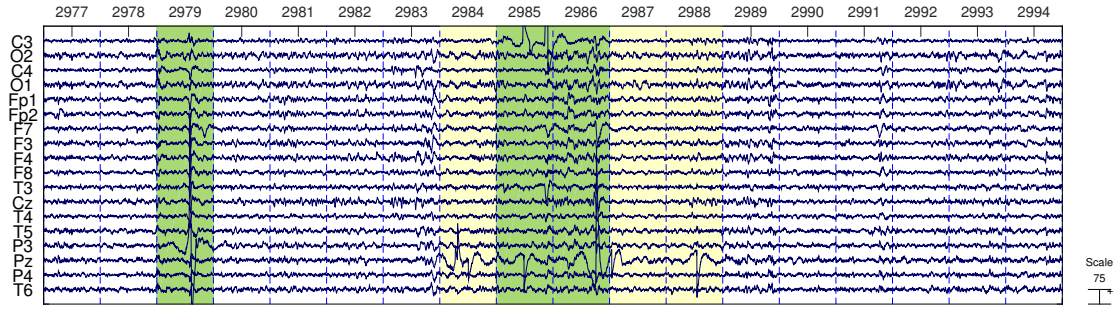
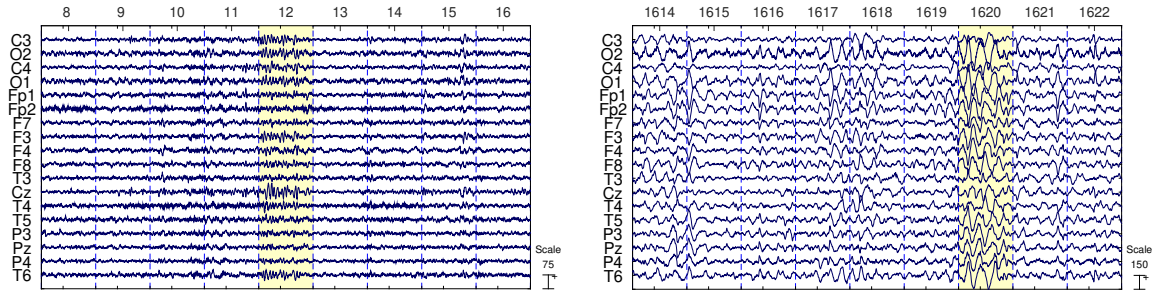


Figure 3.7: False negative detection by FASTER highlighted with yellow color. Correctly detected epochs are labeled with green color. Vertical dashed lines separate 5-s epochs, epoch numbers are given above.

REM 1. It is also a common situation for non-REM 3 where there is a small number of artifacts and there a lot of epochs with high amplitude delta waves. Many epochs with high amplitude delta waves corresponding to non-REM 3 injections and K-complexes are also were labeled as FP. Examples of FP in non-REM 1 and non-REM 3 are depicted in Figure 3.8.

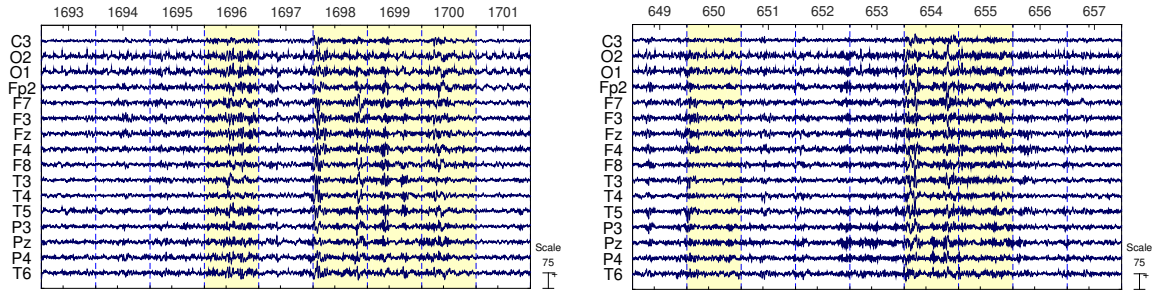
AFAST reduces a feature set calculated in FASTER to analyze abnormal activity in the epoch. It obtains a mean feature value over channels to distinguish all-channels artifacts and maximum relative feature value to distinguish single-channel artifacts. Such an approach allows AFAST to increase sensitivity and, consequently, K. However, the number of false detection increases as well. It detects all epochs with an abnormal combination of median and maximal relative characteristic feature. As it is shown in Figure 3.4 some non-artifact epochs may lie outside the high-density cluster and, consequently, be labeled as an artifact. It often happens in non-REM 2 and 3 due to high-amplitude slow wave normal patterns. In REM, it detects low amplitude eye movement propagation, which was not labeled as an artifact by the expert. Examples are depicted in Figure 3.9. Due to used features, both AFAST and FASTER fail to detect small amplitude fast frequency artifacts in comparison to FASST and SCADM.

A Windows 10 Pro computer with an Intel Core i7 2.40-GHz CPU and 16 GB of RAM was used for testing. We run FASTER, FASST and SCADM using MATLAB R2014b. The average processing time was 8.64 ± 2.54 s, 30 ± 9.6 s and 145.22 ± 44.30 s per recording respectively. The average length of recording is introduced in Table 3.1, the average number of channels is 17 ± 2. The most expensive process of SCADM is extraction of band power information. Testing of AFAST was performed using Python 2.7 and MATLAB R2014b. It took 98.08 ± 32.05 s.



(a) (b)

Figure 3.8: False positive detection by both FASTER and AFAST in non-REM 1 (**a**) and non-REM 3 (**b**) highlighted with yellow color. Vertical dashed lines separate 5-s epochs, epoch numbers are given above.



(a) (b)

Figure 3.9: False positive detection by AFAST in REM (**a**) and non-REM 2 (**b**) highlighted with yellow color. Vertical dashed lines separate 5-s epochs, epoch numbers are given above.

3.4 Discussion

Artifact detection in sleep EEG processing pipeline problem is discussed in this chapter. Even the pipeline assumes the manual artifact rejection, the decision making support will make scoring more consistent and less tedious for the expert. We employ and test several automatic artifact detection strategies, including FASST, SCADM, and FASTER. Moreover, we propose an adaptive improvement of the FASTER method. Testing is performed on data recorded from subjects diagnosed with insomnia. Overall, all the methods able to detect almost all large movement artifacts. However, problems arise in the detection of more complex artifacts and inconsistency with subjective expert judgment may affect the final result.

Testing results show overcome of AFAST over the other methods in terms of agreement with the expert scoring. It outperforms strategies proposed for sleep data analysis. Per-

formance of the method is dependent on data and prevailing types of artifacts. Moreover, expert scoring is a subjective process, especially when there are more than 3500 epochs to label. That explains the poor performance of FASST aimed for pop-up electrode and muscular artifact detection when data contain plenty of channel loss of contact artifacts. However, we believe that increasing recording quality improves the effectiveness of this strategy. The other strategy SCADM was initially proposed for rejection of 20-s epochs of a single channel EEG. Original paper [78] reports sensitivity over 90%. However, scoring rules for 5-s epochs and 20-s epochs may differ. In the first case, scoring is performed in more details, and in terms of no strict rules for artifacts, it generates inconsistency.

Application of FASTER, a popular strategy for wake EEG, showed stable results. It produced a big number of false detection due to sleep EEG specificity. Normal sleep EEG patterns extremely vary in comparison to wake and an increasing threshold for z-scored values do not make any effect. Even using stable statistics in z-scoring procedure does not allow FASTER to overcome this problem. In AFAST, we propose to use clustering instead of thresholding of z-scored values. Obtained FASTER based features reflect outlier epochs by comparison with other epochs and by comparing channels within epochs. Unfortunately, obtained clusters are not spherical. We utilize the density-based clustering method to adapt to a complex cluster form and improve performance results on this dataset. All presented strategies have their advantages, but the important tendency could be noted: the adaptive methods are more preferable for processing of EEG data, especially in terms of high inter-subject variability. Nevertheless, the choice of the artifact detection method chosen for practical application should be made based on data quality and specificity.

3.4.1 Limitations and future work

The main issue of both FASTER and AFAST is that they fail to detect low amplitude fast frequency artifacts. Other strategies (FASST) or additional frequency-specific feature (e.g., number of zero crossings) could be used instead. Another issue is connected to the selected clustering method DBSCAN. First of all, the great computational time is connected with the computation of all-to-all distances, which grows extremely with the increasing number of computed epochs. In the case of one-night recording, data are limited by the natural constraint of sleep length and sleep stage proportions. In case of a significantly longer record, we recommend processing of each sleep recording separately. Moreover, data preprocessing could be applied before the processing by sleep stages. We reject all epochs with an abnormally high average correlation between channels. This condition will be valid only in the case of widely distributed electrodes. For instance, that high-correlated signals recorded from the close locations is expected. Therefore, that

strategy will fail in data recorded that are unevenly distributed over the head electrodes. Moreover, that high-correlated effect could be caused by the chosen montage scheme. Finally, using expert scoring from only one expert also could lead to overfitting. Unfortunately, it was not possible to gain other expert evaluations due to the time required for the scoring of one recording.

Future work will be concentrated on two main aspects. The first one is connected with the investigation of DBSCAN parameters, their connection with signals parameters. DBSCAN has substantial advantages and disadvantages, which caused heated discussion in the scientific society [44], [119]. Therefore, we will try to utilize other suitable clustering methods. However, we prefer simpler approaches to stay more relevant for practice. One of the possibility is to change the feature extraction procedure to get more spherical clusters. The other direction for research is connected to collecting more expert scorings for that data from the same and other experts. We would like to investigate consistency, differences, and uncertainties. That will help to establish more precise criteria for artifact labeling. Moreover, we would like to employ an incremental learning procedure to adapt to the scorer judgment and relieve his work. Finally, we would like to investigate the dependence of artifact-free analysis results (spectral information, properties of important EEG patterns) from artifact rejection procedure, and scoring rules for manual investigation.

3.5 Conclusion

The pipeline for automatic long-term EEG processing is discussed. This pipeline gives insight into the investigation process in the sleep research laboratory. The study tests several strategies for artifact rejection in sleep and wake EEG. Moreover, it proposes an adaptive artifact detection method based on FASTER features. In comparison to FASTER well-known for awake EEG, here, we propose to separate artifact-free epochs using clustering with DBSCAN. This concept allows AFAST to adapt to the data with high inter-variability such as subjects suffering from insomnia. The efficiency of the proposed method was demonstrated in comparison to the original FASTER method and two methods, which were previously proposed for clinical work.

Chapter 4

Artifact Detection Method Using Riemannian Geometry

4.1 Introduction

This chapter presents an unsupervised multichannel approach for artifact detection in sleep EEG based on Riemannian geometry. The aim of the research is to develop an artifact detection method in multichannel sleep EEG capable of rejecting different artifact types at once. The inspiration for the study is gained from the idea of the Riemannian potato [12], an automatic and adaptive artifact detection method proposed for online brain-computer interface (BCI) experiments. It considers spatial patterns of awake EEG data and exploits the fact that clean data tend to form a cluster in the manifold of symmetric positive matrices (the "potato"), the space, to which the EEG data is mapped, while artifacts tend to be distributed outside the cluster. In our study, we assume that the number of clusters for clean data may be superior to one, and we introduce an unsupervised and adaptive method for constructing the clusters. It allows us to achieve better performance for complex data such as sleep EEG, where we cannot assume a unique cluster because the spatial patterns vary widely across sleep stages and because sleep data within the same stage may be captured more accurately by several clusters. Different sleep stages have their own characteristics. In the present study, we investigate the proposed method and test it on real sleep EEG data. The performance is evaluated by comparing the obtained detection to a provided expert scoring. The results are compared to the outcome of the (one-cluster) Riemannian potato artifact detector [12] and the FASTER artifact detection method [93], which performs artifact detection by analyzing statistical properties in separate channels.

In comparison to the methods proposed in the previous chapter, it provides continuous evaluation of the data instead of detection contaminated epochs of the same length. That

leads to more accurate detection and, consequently, allows saving more artifact-free EEG data for further analysis. The text of this chapter is an extended version of the paper: E. Saifutdinova, M. Congedo, D. Dudysova, L. Lhotska, J. Koprivova, and V. Gerla, "An unsupervised multichannel artifact detection method for sleep eeg based on riemannian geometry", *Sensors*, vol. 19, no. 3, 2019, ISSN: 1424-8220. DOI: 10.3390/s19030602.

4.2 Materials and methods

4.2.1 Data

Testing of the methods required data with a detailed artifact scoring. Therefore, using data from the previous chapter is not suitable. Here we used two datasets described in following sections, which consist of short-term data recordings with detailed artifact scoring.

Dreams dataset

The open-source Dreams dataset [32] contains 15-min recordings of sleep data, 20 recordings in total. Channels originally referenced to A1 were re-referenced to Cz. All available EEG channels O1-Cz, O2-Cz, Fp1-Cz, and Fp2-Cz were used in the study. The sampling frequencies were 50 Hz, 100 Hz, and 200 Hz. We excluded from the analysis the recordings sampled at 50 Hz due to the fact that with such a sampling rate, the frequency band-pass is too restricted for sleep data analysis [56]. Furthermore, we excluded recording Numbers 3, 6, 10, and 16 because the proportion of artifacts exceeded 40%. Sleep stage scoring and artifact labeling were performed by a trained clinician and provided, as well. Each recording contains a mix of sleep stages. The distribution of the sleep stages in the dataset is: wakefulness 33%, REM 25%, non-REM 1 18%, non-REM 2 23%, and other 1%. Information on the artifact rate is shown in Table 4.1.

InSleep datasets

The InSleep datasets were obtained from the National Institute of Mental Health, Czech Republic. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of the National Institute of Mental Health (Project Number 6/15). The whole-night EEG as a part of the polysomnography (PSG) was recorded from subjects suffering from insomnia. Each recording was obtained using 19 channels placed according to the 10-20 system: Pz, Cz, Fz, T6, T5, T4, T3, F8, F7, O2, O1, P4, P3, C4, C3, F4, F3, Fp2, Fp1. Channels were referenced to an average of Cz, Fz, and Pz. The sampling frequency was 250 Hz. Sleep stages were scored by a trained clinician. Each recording contained 10–20 min of data corresponding to a single

sleep stage (REM, non-REM 2, or non-REM 3) or awake activity. The non-REM 1 sleep stage was not present in the dataset due to its short duration in sleep recordings. Artifacts were labeled by an expert. Visual inspection of EEG was performed separately for each recording in the EEG Lab software using function "reject continuous data by eye" [30]. The expert was instructed to label all visible artifacts. The total number of obtained recordings was 44. Details of the artifact rate are provided in Table 4.1.

Table 4.1: Datasets details.

Dataset	**Sleep Stage**	**Number**	**Length (min)**	**Artifact Rate (%)**
Dreams		12	15	9.38 ± 9.84
InSleep	Wakefulness	14	15.38 ± 2.26	18 ± 9.9
	REM	12	13.5 ± 1.88	7.43 ± 5.82
	Non-REM 2	10	15.3 ± 1.77	4.07 ± 5.33
	Non-REM 3	8	14.75 ± 2.76	1.97 ± 1.43

4.2.2 Theoretical background

This section describes the theoretical aspects of working with multichannel data. We denote by $x_t \in R^n$ the signal vector at n channels and at time point t. Then, $X = [x_s, ..., x_{s+m-1}] \in R^{n \times m}$ denotes an epoch starting at sample s and lasting m samples. For X, we estimate the spatial covariance matrix using the usual sample covariance matrix (SCM) estimator:

$$C = \frac{1}{m-1} X X^T \in R^{n \times n}. \tag{4.1}$$

The matrix C is symmetric positive-definite (SPD). Often, SPD matrices are considered in a Euclidean space with the associated Frobenius norm $\|C\|_F$. However, the native space of SPD matrices is not Euclidean, but a Riemannian space [29]. Riemannian geometry starts by defining an inner product (metric) at each tangent space to the manifold, which varies smoothly from point to point. Applying the affine-invariant (Fisher) metric [29], the distance δ between two points C_i and C_j on the manifold is given by:

$$\delta(C_i, C_j) = \left[\sum_{i=1}^{n} log^2 \lambda_i \right]^{1/2}, \tag{4.2}$$

where the λ_i are the eigenvalues of $C_i^{-1} C_j$.

A centroid (called the geometric mean) of an SPD matrix set can be defined using the above distance [26]. The geometric mean M of a covariance matrix set C_i is calculated using an iterative algorithm. First, M is initialized by the arithmetic mean of C_i. Then,

the following iteration is repeated:

$$M = M^{1/2} \exp[1/k \sum_k \ln(M^{-1/2} C_k M^{-1/2})] M^{1/2} \tag{4.3}$$

until convergence, which is obtained when:

$$\| \sum_k \ln(M^{-1/2} C_k M^{-1/2}) \|_F < \varepsilon. \tag{4.4}$$

The distance of a set of covariance matrices to their geometric mean does not have a symmetric distribution. To make the distribution symmetric, we may use the standardized distances to the geometric mean [28]; define the geometric mean μ of the distances δ_k as:

$$\mu = \exp(\frac{1}{k} \sum_k \ln \delta_k) \tag{4.5}$$

and the geometric standard deviation σ of the distances as:

$$\sigma = \exp(\sqrt{\frac{1}{k} \sum_k (\ln \frac{\delta_k}{\mu})^2}). \tag{4.6}$$

Then, the standardized distance measures are:

$$\delta'_k = \frac{\ln \frac{\delta_k}{\mu}}{\ln \sigma}. \tag{4.7}$$

There are known applications of Riemannian geometry in biomedical signal processing for sleep stage classification [71] and detection of respiratory states [90].

4.2.3 Riemannian potato

The Riemannian potato artifact detector (RP) approach [12] is based on a simple idea that contaminated epoch differs enough from the clean signal. It uses covariance matrix as a signal descriptor. Defining a set of examples it estimates average covariance matrix estimated on the signal baseline as a geometric mean. Then it computes z-scored distribution of distances between covariance matrices and obtained geometric mean of baseline data as it is shown in Figure 4.1. Artifacts are detected as covariance matrices beyond the region limited by a predefined z-score value. It is called potato due to its crooked aspect, which is induced by the non-linearity of the Riemannian manifold [13]. Artifact rejection is mainly used for tasks like in P300-based BCI spellers [81] and games [9]. There exist application in medical EEG for cognitive studies [124] and psychiatric disorder [7].

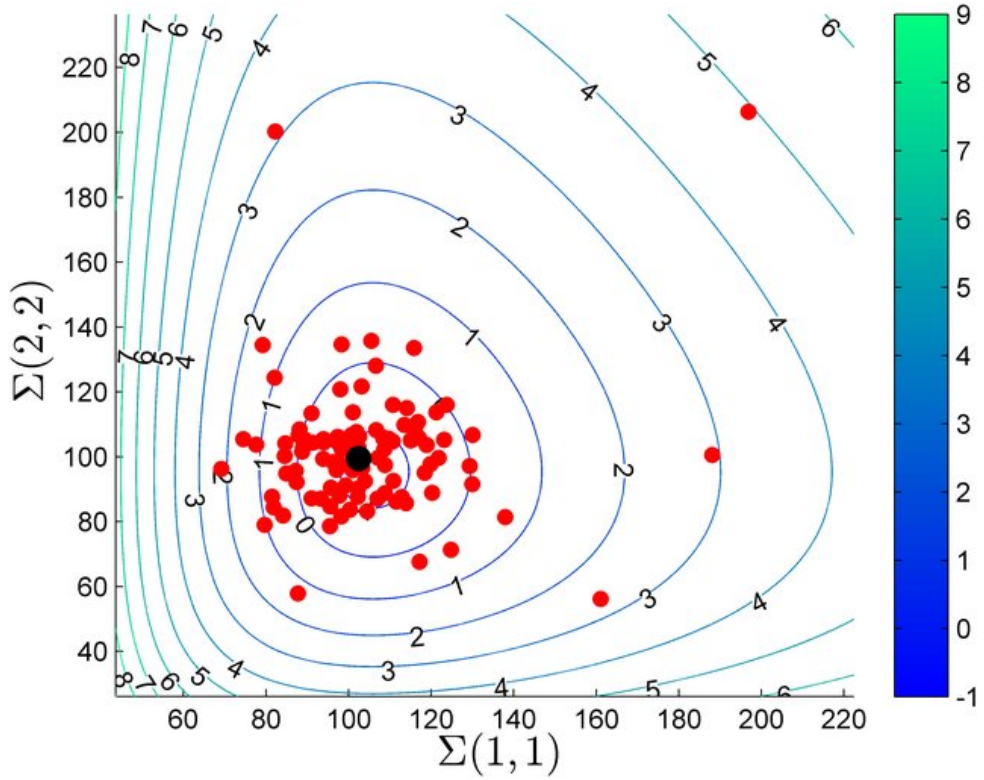


Figure 4.1: 2D projection of the z-score map of a Riemannian potato from [13]. There are 100 simulated 2×2 matrices Σ (in red) and their reference matrix (in black). The colormap defines the z-score and a chosen isocontour z_{th} defines the potato.

4.2.4 The proposed method: RPs

A Riemannian potatoes artifact detection method (RPs) is proposed in this study. It relies on the assumption that on the SPD manifold, contaminated epochs are mapped far away from clusters of clean data. In contrast to the simple potato [12], here we take into consideration the fact that several clusters may be needed to describe the distribution of clean sleep data. In the first step, RPs restore these clusters from a recorded EEG. In the second step, we obtain continuous scoring by analyzing the data in a sliding window and comparing them to the obtained clusters. Finally, artifacts are detected by applying a threshold. The method overview is depicted in Figure 4.2. As in [12], we filtered the data before further processing. However, we used a low-pass filter under 30 Hz to save frequencies significant for sleep stages [56].

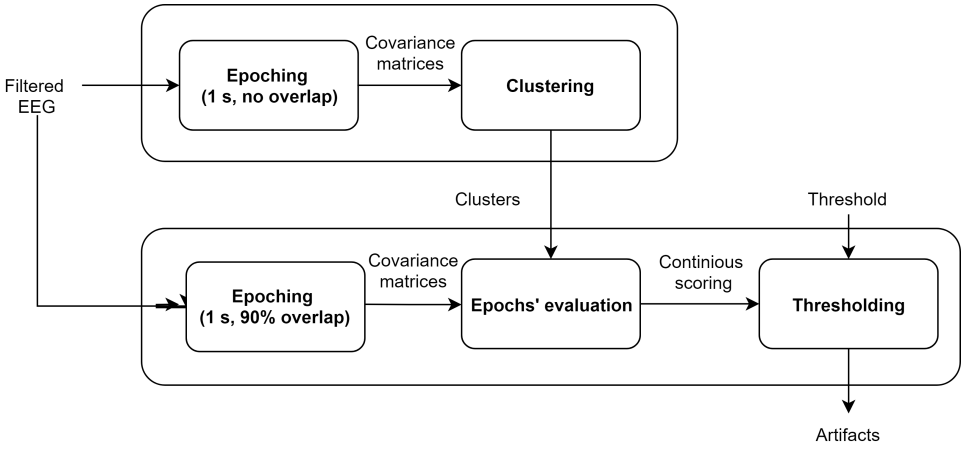


Figure 4.2: The method overview.

The cluster construction process starts with splitting the filtered data into non-overlapping 1-s epochs. Such an epoch length is a standard for Riemannian geometry analysis [12]. Covariance matrices are extracted from every epoch. Then, distances between all pairs of covariance matrices are calculated by Equation (4.2). The average distance d_i is the mean value of distances from the ith epoch to all the other epochs. Epochs with a d value greater than a threshold are removed from further analysis. We consider the mean value of d as the threshold. Then, we cluster the remaining matrices using the k-means algorithm with the distance measure defined by Equation (4.2). To determine the number of clusters, we use the following procedure: the number of clusters increases iteratively from one to a maximum of 10; centroids are obtained as the geometric means; the process is finished once the constructed clusters satisfy an overall normality condition, that is when the distribution of standardized distances to the centroids is normal in all clusters. Normality is tested using D'Agostino's K^2 test. The distribution of distances to the centroid within a cluster is examined separately. The overall normality p-value is obtained as a combination of the obtained p-values using Stouffer's Z-score method. The cluster construction process is stopped when the combined p-value is superior to 0.05. If the cluster construction finishes without achieving overall normality, then we choose the number of clusters achieving the greatest p-value. In the case of tied p-values, we retain the one corresponding to the smaller number of clusters. The rationale for this procedure is that when the distances of the points in the cluster from the center of mass have a Gaussian distribution, we assume that the cluster is a good representative of a region in the manifold covered by the data and that further splitting is therefore unnecessary.

In the second step, RPs analyze the n-channel data in a sliding window of a size of 1 s with a step equal to 0.1-times the sampling frequency. Assume there are k clusters obtained in the previous step and denote $O_j \in R^{n \times n}, j \in [1, k]$ their centroids. For each ith window, we obtain covariance matrix $C_i \in R^{n \times n}$. All k distances σ_{ij} between C_i and

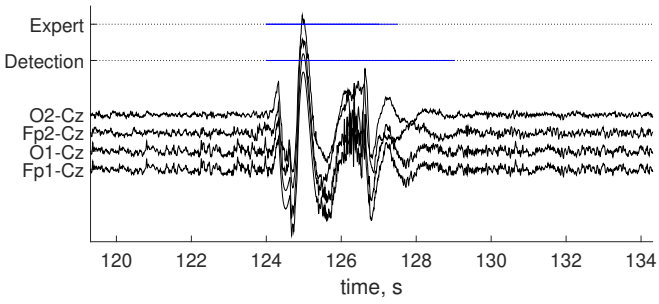
every O_j centroid are calculated. Then, the closest cluster j^* to C_i is determined by the minimum distance among the σ_{ij} values. The geometric mean and standard deviation within the cluster were determined in the cluster construction step. Therefore, we can obtain σ_i^* as a standardized distance σ'_{ij^*} to the centroid O_{j^*} by Equation (4.7). Due to the cluster construction, standardized distances within a cluster are distributed normally. We can consider the obtained σ_i^* as a z-score and translate it into a probability value p_i using the Gaussian cumulative distribution function. The value p_i indicates a probability of epoch i with covariance matrix C_i to belong to a given cluster. Hence, the value of p'_i equals $1 - p_i$ and indicates a probability of epoch i with covariance matrix C_i to be an outlier.

We consider that the values σ_i^* correspond to the center of the ith epoch and that in an epoch, the scoring starts and ends with 0. Therefore, we interpolate the obtained values within the epoch. The obtained score is then smoothed with a moving average filter to make the output more robust. The local minima of these scores define segments. Finally, a segment whose score is above a threshold for at least 0.4 s is considered as an artifact. This value was established as a minimum artifact duration by the experts.

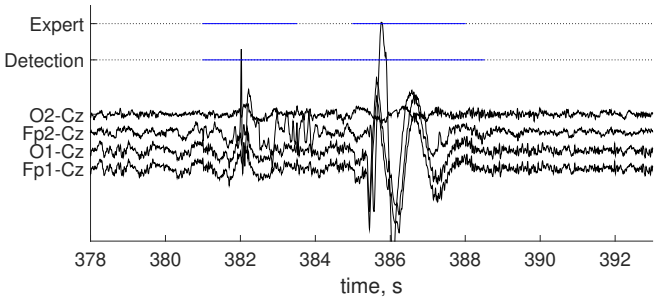
4.3 Experiment results

The proposed RPs artifact detection method was implemented in Python and applied to the Dreams and InSleep datasets. The Riemannian potato [12] method was chosen as the benchmark. We used the original implementation of the Riemannian potato artifact detector provided by the authors in the covariance toolbox for MATLAB (`https://github.com/alexandrebarachant/covariancetoolbox`). Performance evaluation utilizes *Strategy 2* from Section 2.4.4 to discover ability of the method to separate data signals in two classes. However, precise artifact borders are hard to determine. Human and automatic scoring may differ and yet be both considered correct, as shown in Figure 4.3. To handle this in a sample-by-sample evaluation, we added a fuzzy area on each artifact border labeled by an expert. Thus, residual segments with no matching labeling as segments A-B and C-D in Figure 4.4 were treated as TP and TN, respectively, if they lasted for at least 10% of the artifact duration and no more than 1.5 s. Otherwise, they were counted as FP or FN, respectively. The parameters used for results evaluation were estimated based on experience and consultancy with neurological experts who described their visual inspection process. Cohen's kappa K, sensitivity Se, and false discovery ratio FDR were calculated regarding presented epoch evaluation.

The non-parametric Wilcoxon signed rank test was used for comparing metrics. This test does not require normality of the samples and is also suitable for small sample sizes.



(a)



(b)

Figure 4.3: Examples of comparison of human scoring and artifact detection (**a**,**b**).

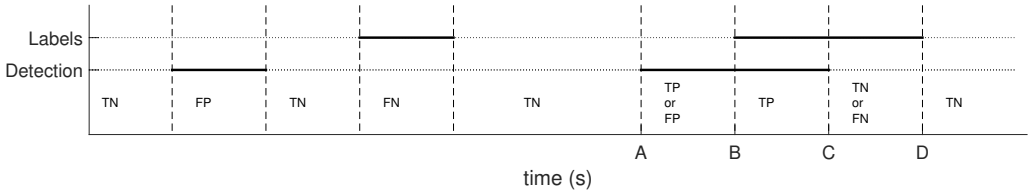


Figure 4.4: Detection evaluation. Horizontal bold lines represent detected segments. The horizontal dotted line stands for segments labeled as artifact-free data. Vertical dashed lines are the ends of all labeled segments.

Visual comparison of scoring obtained by RPs and by the benchmark is provided in Figures 4.5 and 4.6. Detection results were compared with scoring provided by an expert, and the three statistical metrics presented in the previous section were computed as well as percentage of data labeled as artifacts by methods. Each recording of the Dream dataset contains a mix of sleep stages and awake EEG. Some of the presented sleep stages have a short duration and do not include artifacts. Testing results grouped for such EEG types would not be representative. Therefore, metric values for the Dreams dataset were obtained for the entire recordings. Recordings in the InSleep dataset contain activity of a single stage. That allowed us to assess the method performance on separate sleep stages and wakefulness. Table 4.2 reports the obtained results for the benchmark, and Table 4.3 contains values for the proposed method. Cohen's kappa for both methods on all datasets is shown in Figure 4.7.

The proposed method, RPs, performed better than the benchmark in terms of FDR

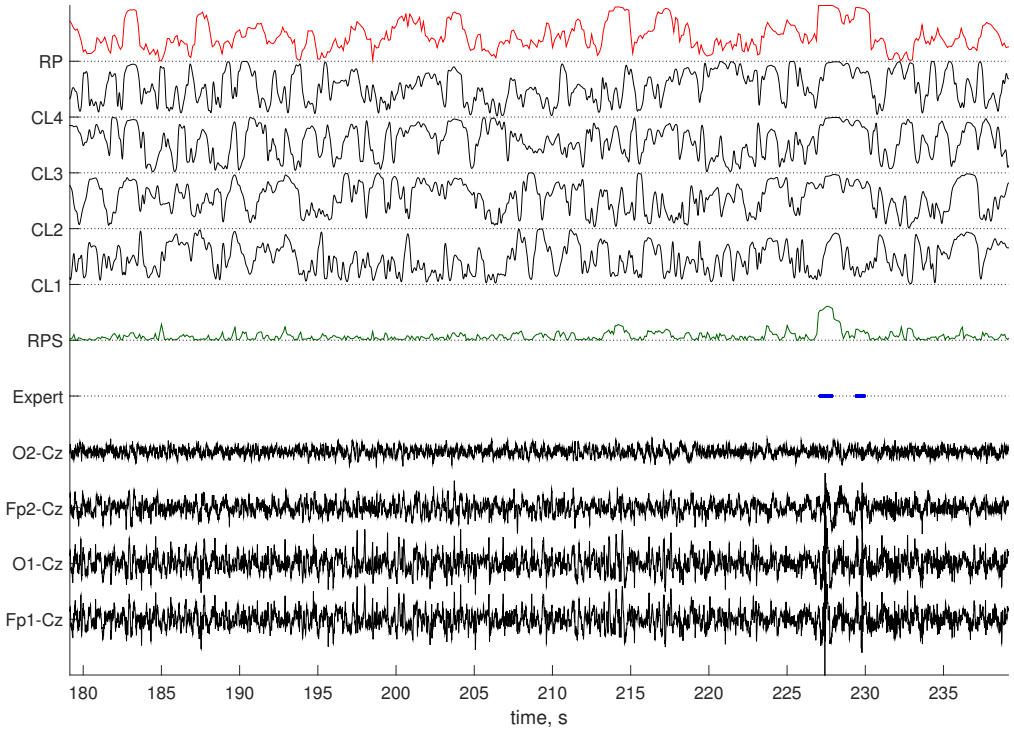


Figure 4.5: Scoring obtained by the benchmark method (RP) and RPs (RPS) on data from the Dreams dataset. Normalized distances to four clusters obtained for this recording with RPs (CL1–4) are shown, as well. Artifacts (Expert) are denoted with the solid blue line.

Table 4.2: Performance of the benchmark artifact detection method.

Dataset	**Sleep Stage**	$\boldsymbol{K}$	$\boldsymbol{Se}$	$\boldsymbol{FDR}$	**Artifact Rate (%)**
Dreams		0.58 ± 0.2	0.79 ± 0.19	0.4 ± 0.21	12.23 ± 13.71
InSleep	Wakefulness	0.84 ± 0.09	0.91 ± 0.08	0.16 ± 0.07	15.2 ± 8.24
	REM	0.75 ± 0.24	0.88 ± 0.12	0.28 ± 0.24	8.3 ± 7.14
	Non-REM 2	0.40 ± 0.24	0.77 ± 0.21	0.62 ± 0.41	9.12 ± 13.53
	Non-REM 3	0.54 ± 0.17	0.62 ± 0.29	0.37 ± 0.21	1.58 ± 1.52

Table 4.3: Performance of RPs artifact detection.

Dataset	**Sleep Stage**	$\boldsymbol{K}$	$\boldsymbol{Se}$	$\boldsymbol{FDR}$	**Artifact Rate (%)**
Dreams		0.74 ± 0.17	0.78 ± 0.18	0.24 ± 0.22	7.49 ± 7.65
InSleep	Wakefulness	0.85 ± 0.09	0.9 ± 0.08	0.12 ± 0.08	16.6 ± 7.62
	REM	0.82 ± 0.11	0.86 ± 0.12	0.18 ± 0.13	9.47 ± 8.22
	Non-REM 2	0.50 ± 0.37	0.84 ± 0.17	0.55 ± 0.38	9.06 ± 5.81
	Non-REM 3	0.68 ± 0.22	0.68 ± 0.26	0.22 ± 0.22	1.42 ± 1.5

for the Dreams dataset (p-value < 0.05) and InSleep REM and non-REM 2 data (p-value < 0.05 and p-value < 0.01, respectively). With respect to sensitivity, RPs did not outperform the benchmark. Finally, RPs significantly outperformed the benchmark method on non-REM 2 (p-value < 0.01) and non-REM 3 (p-value < 0.05) in terms of agreement with the expert scoring. A similar trend is observed for the InSleep dataset, as shown in Figure 4.7 (p-value < 0.1).

Visual inspection in the Dreams dataset revealed that the benchmark method tended to detect incorrectly as artifacts characteristic EEG patterns of non-dominant sleep stages such as high amplitude delta-waves corresponding to non-REM 3 activity in recordings with a dominant non-REM 2 sleep stage, as illustrated in Figure 4.8.

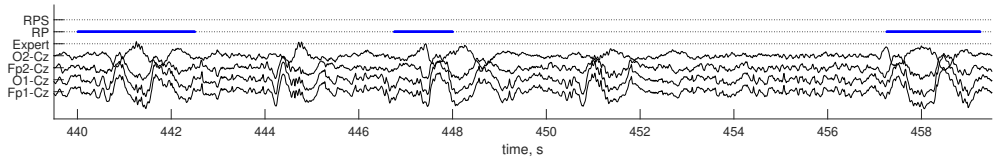


Figure 4.8: Dreams dataset. Lines represent the expert scoring (Expert), results of the benchmark (RP), and the proposed (RPS) methods.

FDR values obtained on InSleep non-REM 2 indicated a big proportion of false positive events. Both methods incorrectly detected injections of non-REM 3 sleep in non-REM 2. This activity is defined as high amplitude slow waves lasting 1–5 s. An example is provided in Figure 4.9a. Such events account for a small proportion of the recording, for which the methods do not appear adapted. Furthermore, the non-REM 2 sleep data also contained K-complexes and sleep spindles characterized as a single high-amplitude delta wave and sigma bursts, respectively. An example is depicted in Figure 4.9b. A wide range of patterns of such events complicates the detection for both methods. Moreover, artifacts in stages non-REM 2 and non-REM 3 occurred rarely according to Table 4.1; this made the metrics values for these datasets more sensitive to false detections.

As we have mentioned, RPs utilizes 1-s epochs, which is a standard window length for Riemannian geometry analysis. We have also applied RPs using 2-s and 4-s epoch lengths to all datasets to test whether longer epoch lengths are more suitable for capturing the dynamics of slow waves. The obtained results are presented in Tables 4.4 and 4.5 respectively. The results showed no significant increase in K in all datasets. In fact, a significant decrease in K was observed for wakefulness and REM InSleep (p-value < 0.01). Indeed, the number of events corresponding to normal slow wave activity and incorrectly determined as artifacts in non-REM 2 and 3 decreased. However, using such modifications decreased the number of detected events in general and increased the number of false results. This is caused by an increased number of missed small artifacts. Moreover, the

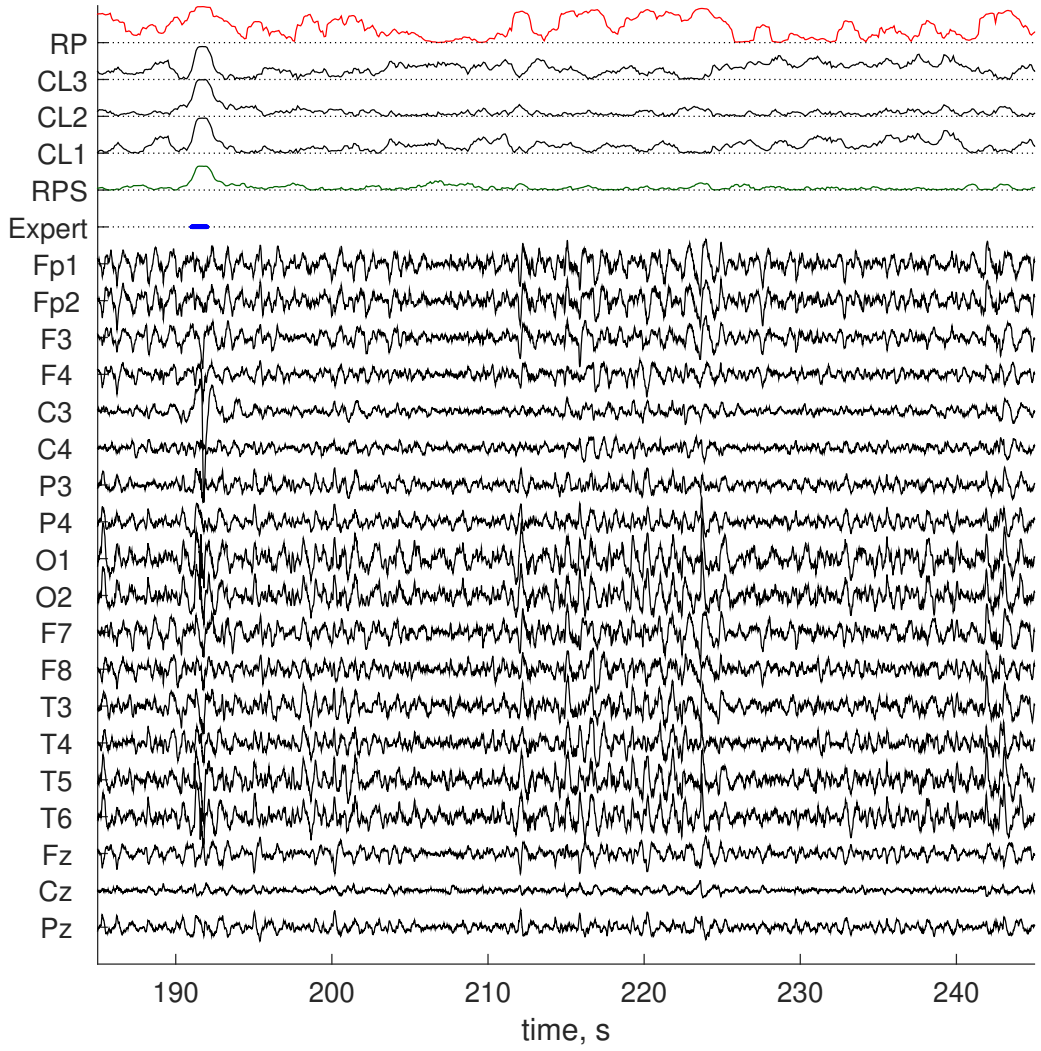


Figure 4.6: Scoring obtained by the benchmark method (RP) and RPs (RPS) on data from the InSleep dataset. Normalized distances to three clusters obtained for this recording with RPs (CL1–3) are shown, as well. Artifacts (Expert) are denoted with a solid blue line.

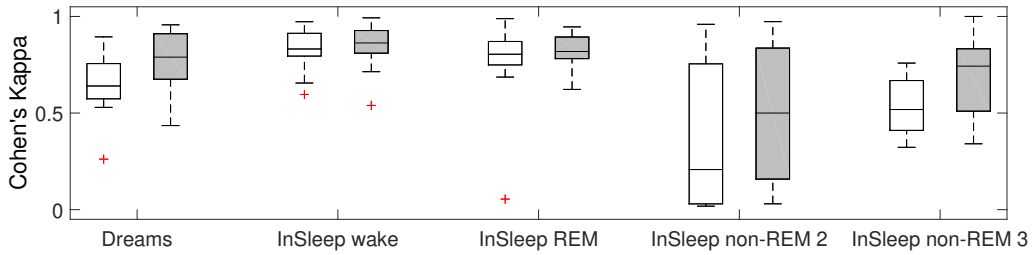


Figure 4.7: Distribution of Cohen's kappa for the Dreams and InSleep datasets. White boxes represent the benchmark; grey boxes represent the proposed method.

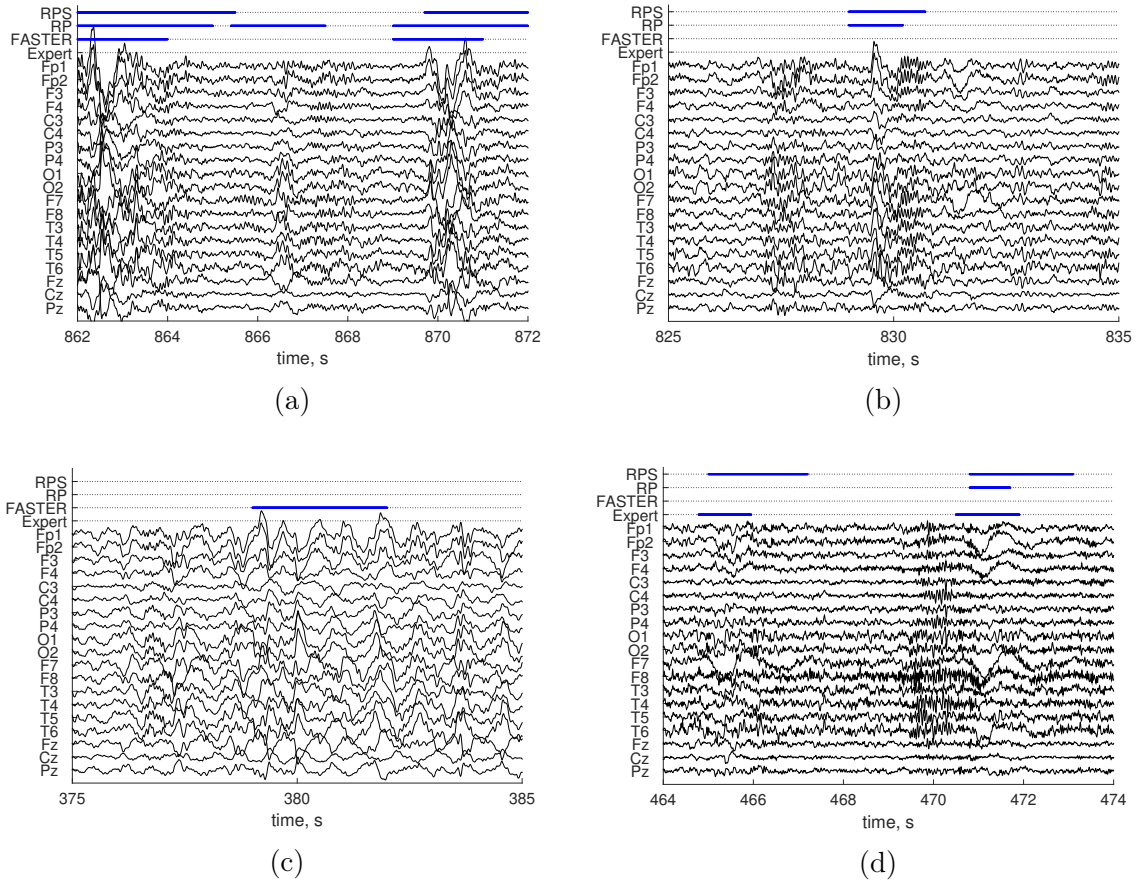


Figure 4.9: Expert scoring (Expert), results of the FASTER (FASTER), benchmark (RP), and the proposed (RPS) methods of different EEG patterns in delta waves (**a**), K-complex in non-REM 2 (**b**), delta waves in non-REM 3 (**c**), and ocular artifact in REM (**d**).

detected events were too long, which led to an increase of the FP value and, consequently, FDR.

Additionally, we have applied FASTER [93] for a window size 1 s to all datasets. This method rejects epochs of a single-channel EEG based on calculated thresholds. The process is repeated for each channel. Mean, variance, amplitude range, and mean gradient are used for decision making. The results are presented in Table 4.6. The proposed RPs method significantly outperformed FASTER in K for all datasets (p-value < 0.05 for InSleep non-REM 3, p-value < 0.01 for others). Moreover, RPs achieves significantly better Se in InSleep wakefulness (p-value < 0.01), REM, and non-REM 2 (p-value < 0.05 for both datasets) and significantly smaller FDR for Dreams, InSleep non-REM 2 and 3 (p-value < 0.01, p-value < 0.05, and p-value < 0.01, respectively). The method performed well in the detection of movement artifacts. However, often, it incorrectly detected a high amplitude delta in non-REM 2 (Figure 4.9a) and non-REM 3 (Figure 4.9c) as an artifact. Furthermore, it missed many short and small-amplitude artifacts (Figure 4.9d).

Table 4.4: Performance of RPs artifact detection with a 2-s epoch length.

Dataset	Sleep Stage	K	Se	FDR	Artifact Rate (%)
Dreams		0.71 ± 0.17	0.85 ± 0.1	0.31 ± 0.17	12.17 ± 13.29
InSleep	Wakefulness	0.77 ± 0.13	0.85 ± 0.1	0.2 ± 0.14	18.75 ± 8.17
	REM	0.72 ± 0.14	0.8 ± 0.19	0.25 ± 0.19	8.99 ± 5.31
	Non-REM 2	0.53 ± 0.41	0.79 ± 0.21	0.49 ± 0.44	9.75 ± 7.78
	Non-REM 3	0.67 ± 0.26	0.67 ± 0.26	0.26 ± 0.27	2.74 ± 2.08

Table 4.5: Performance of RPs artifact detection with a 4-s epoch length.

Dataset	Sleep Stage	K	Se	FDR	Artifact Rate (%)
Dreams		0.52 ± 0.16	0.8 ± 0.14	0.54 ± 0.16	15.12 ± 15.27
InSleep	Wakefulness	0.68 ± 0.11	0.84 ± 0.1	0.32 ± 0.14	20.75 ± 8.62
	REM	0.53 ± 0.16	0.66 ± 0.2	0.46 ± 0.2	9.59 ± 7.92
	Non-REM 2	0.46 ± 0.38	0.77 ± 0.19	0.57 ± 0.38	9.52 ± 7.74
	Non-REM 3	0.55 ± 0.23	0.62 ± 0.3	0.46 ± 0.24	2.87 ± 2.18

Table 4.6: Performance of the FASTER artifact detection method.

Dataset	Sleep Stage	K	Se	FDR	Artifact Rate (%)
Dreams		0.31 ± 0.22	0.81 ± 0.28	0.66 ± 0.26	24.9 ± 11.55
InSleep	Wakefulness	0.58 ± 0.21	0.52 ± 0.24	0.09 ± 0.14	6.45 ± 6.14
	REM	0.46 ± 0.21	0.57 ± 0.24	0.45 ± 0.27	8.64 ± 6.3
	Non-REM 2	0.28 ± 0.25	0.43 ± 0.36	0.72 ± 0.25	3.46 ± 2.12
	Non-REM 3	0.41 ± 0.21	0.66 ± 0.34	0.64 ± 0.16	3.29 ± 1.57

For the estimation of time consumption, we ran FASTER and RPs on 15-min 19-channel recordings. A Windows 10 Pro computer with an Intel Core i7 2.40-GHz CPU and 16 GB of RAM was used. Testing of FASTER, excluding the ICA decomposition, which was very time consuming, was performed using MATLAB R2014b. The processing time was 5.44 ± 0.01 s per recording. Testing of RPs was performed using Python 2.7. The processing time of RPs was 43.86–112.63 s per recording depending on the number of constructed clusters. Cluster construction took 0.03 s for a single cluster and 68.72 s for the construction of six clusters. Testing on the Dreams dataset resulted in 1–4 clusters. The number of clusters in the InSleep dataset varied in the range of 1–6 for awake and REM EEG and in the range of 1–4 for non-REM 2 and 3 activity. The number of clusters did not depend on the number of channels, but only on the presence of different activities, which could be distinguished in terms of the Riemannian geometry of the data.

4.4 Discussion

The performance of RPs was compared to a previously-developed artifact detection method based on Riemannian geometry. Testing on rest-state awake data confirmed the effectiveness of both RPs and the benchmark approaches. The results obtained for sleep data proved that the proposed method was more favorable. The reason why RPs outperformed the benchmark in sleep data is that the multiple cluster construction was adapted to the variability of normal spatial patterns in sleep stages. This resulted in a reduction of false detections and improved agreement with the expert. The method was fully unsupervised and adaptive; therefore, detection was dependent only on the input data and could be completely automated. Testing results proved that on the data we considered, a 1-s epoch length was an optimal choice for RPs. Furthermore, the proposed method is scalable and may be applied for any number of channels. Nevertheless, using highly-correlated signals makes the application of the method cumbersome because the estimated covariance matrices may be badly conditioned. Therefore, in the case of a very large number of EEG channels, a dimensionality-reduction pre-processing step is recommended. To do this, methods inspired by the Riemannian geometry of SPD matrices may be used; see [27], [106].

To employ an epoch-based artifact detection strategy, we set epoch window to 1 s. Among all strategies described in the previous chapter only FASTER and FASST can be easily adopted to 1-s epochs. During testing, FASST showed poor performance due to detected artifact types limitation. Only few artifacts over the whole dataset were detected. In comparison to FASTER, which relies on statistical properties in separate channels, the methods based on Riemannian geometry achieved better performance on all datasets. It is based on statistics and could be highly affected by the short data duration. FASTER missed too many artifacts, mostly small ocular and short electrode artifacts. Nevertheless, all tested methods equally performed well on the detection of high frequency/amplitude artifacts. Such artifacts prevailed in wakefulness, which allowed FASTER to achieve better performance among all datasets therein. For sleep, FASTER overestimated artifact activity, which led to low performance. In non-REM 2, it detected only injections of non-REM 3 activity, which noticeably affected the metric values. All methods detected injections of non-REM 3 sleep into non-REM 2 as an artifact. Even if such events were not labeled by the expert as artifacts, their elimination led to improvement of the spectral characteristics of non-REM 2 activity. As for computational efforts, FASTER significantly outperformed RPs, since it utilized simple signal statistics, whereas RPs was based on eigenvalue estimations for computing Riemannian distances. The time required by the RPs method, however, is low in absolute terms and very acceptable for any practical purpose.

This study shows once more that automatic artifact detection in sleep EEG is a challenging task and demonstrates the low efficiency of methods developed for awake EEG. Many studies addressing this problem proposed methods for artifact detection without information about sleep stages such as FASST [128]. However, EEG investigation pipelines in modern sleep research laboratories include analysis of sleep data with already labeled sleep stages. Some recent studies such as SCADM have proposed artifact detection based on analysis of 20- and 30-s epochs [78]. Evaluation using 20-s (30-s) epochs is required for data investigation at a macro level in sleep scoring by humans, whereas for an automatic artifact rejection, it can be performed at a micro-level, and it is preferable to consider smaller epochs. A way to make the automatic method more comparable to human scoring would be to reject whole 20-s (30-s) segments if the majority of 1-s epochs forming them were rejected. Moreover, sleep disorders increase the inter-subject variability of the EEG, as well as the number of artifacts, which makes the task of artifact detection even more demanding. For these and other situations, a fully-unsupervised and adaptive strategy such as RPs is theoretically appealing.

4.4.1 Limitations and future work

The RPs method we have proposed has several disadvantages and therefore can be further improved. First, infrequent EEG patterns might not be represented in the clusters of artifact-free activity due to the cluster construction procedure. Thus, they might be incorrectly detected as artifacts. An improvement could be achieved by setting an automatic method for identification of expected patterns like K-complexes and post-processing analysis. Moreover, the method seems to fail in rejecting artifacts that are not widely distributed on the scalp, such as muscular or cardiac artifacts. This may be obviated by constructing smaller potatoes that include only a small number of channels and by combining the rejection of all the potatoes (Riemannian potato fields [13]), which is currently under investigation. The method was not tested on high-density EEG. In this case, the problem of highly-correlated signals in high-density EEG must be carefully considered. Future work will also be focused on improvements of the method in order to eliminate false detections of K-complexes and sleep spindles. The application of the method is currently under consideration for semiautomatic artifact rejection at the National Institute of Mental Health of the Czech Republic. At this stage, probability scoring will be provided along with data to support manual investigation. Additionally, the method will be extended to multimodal PSG recordings. Other directions of future research include application of the proposed method and analysis of the obtained clusters for automatic sleep stage identification using supervised machine learning approaches.

4.5 Conclusions

The study presents an unsupervised multichannel artifact detection method for sleep EEG based on Riemannian geometry. The presented method forms clusters of artifact-free EEG data considering their spatial patterns. Application of the proposed method to sleep recordings brings significant benefits. The method is scalable, fully unsupervised and adaptive, independent on artifact types, and the outcome is easily interpreted. In comparison to the Riemannian potato artifact detector, it demonstrates better performance on complex sleep data in terms of agreement with human scoring and reduces the number of events incorrectly identified as artifacts. The RPs toolbox is available at `https://gitlab.ciirc.cvut.cz/open-source/rps`.

Chapter 5

Sleep Spindle Detection Using Adaptive Segmentation

5.1 Introduction

The aim of the research described in this chapter is to propose a novel approach for automatic sleep spindle detection in sleep EEG based on adaptive segmentation. Despite of the fact that we use it for a specific sleep EEG pattern, this method could be adapted for detection of other EEG patterns and, moreover, other domain channels. Such approach may be considered for detection of any pattern determined by shape, frequency, and amplitude.

In the presented approach, we combine the two greatest achievements proposed in previous works to make detection more stable and effective. First, we use a function which detects spindle related information. That idea underlies RSP, RPS, SIGMA, TEO, and other methods in clinical practice. However, here, we employ adaptive segmentation based on spindle related information to obtain event candidates for future processing. This step allows us to reduce the data and leave only those segments, which show enough sleep spindle related information. Then, we separate the obtained set of candidates by amplitude. Instead of using a predefined threshold, we obtain one value from the data using the statistical properties of the set. Here, we apply a GMM firstly introduced by Patti *et al.* [97] in 2014. This allows the method to adapt to the data and finds more sleep spindles labeled by an expert. Besides the performance improvement, the proposed method has another advantage. The idea of the method is simple enough to be used in clinical studies. Often, researchers in sleep medicine prefer to use more conservative algorithms due to their simplicity and transparency. In the proposed method, we enhance segmentation and threshold definition steps to make them more suitable for the data. It is still simple in comparison to complex systems optimized to a specific dataset.

Some thoughts described in this chapter and very first tests are presented in the papers: E. Saifutdinova, M. Macaš, V. Gerla, and L. Lhotská, "Adaptive segmentation optimization for sleep spindle detector", in *Information Technology in Bio- and Medical Informatics*, M. E. Renda, M. Bursa, A. Holzinger, and S. Khuri, Eds., Cham: Springer International Publishing, 2016, pp. 85–96, ISBN: 978-3-319-43949-5 and E. Saifutdinova, V. Gerla, L. Lhotska, J. Koprivova, and P. Sos, "Sleep spindles detection using empirical mode decomposition", in *Computational Intelligence for Multimedia Understanding (IWCIM), 2015 International Workshop on*, Oct. 2015, pp. 1–5.

5.2 Materials and methods

5.2.1 Data

Three datasets were used in this study. First two datasets are obtained from the open-source Dreams database [31]. It has 8 sleep EEG recordings on channel CZ-A1 or C3-A1 of 30 min length. Subjects were diagnozed with different pathologies including dysomnia, restless legs syndrome, insomnia, apnoea/hypopnoea syndrome. Data of six subjects were recorded at 200 Hz and the two others at 50 Hz and 100 Hz respectively. Sleep spindles were labeled by two experts. First expert evaluated all 8 recordings and the second one first six. For simplicity, we refer to these data as two different datasets Dreams V1 and Dreams V2. We select only sleep stage 2 part of the analysis due to consistency in statistical properties of spindle parameters in the data. Also, we excluded recording with frequency sample 50 Hz because of too restricted frequency band-pass [56]. Details on Dreams V1 and Dreams V2 are presented in Table 5.1.

The third dataset NIMH was obtained from the National Institute of Mental Health, Czech Republic. The whole-night EEG as a part of the PSG was recorded from subjects suffering from insomnia. Sleep stages scoring was performed by a trained clinician using [56]. Artifact rejection was performed in 5 second epochs in non-REM 2 sleep using pipeline presented in previous section and additionally controlled by the expert. Sleep spindle labels were scored on C3-A2 EEG channel in sleep stage 2 by an expert. Details on NIMH dataset are presented in Table 5.1.

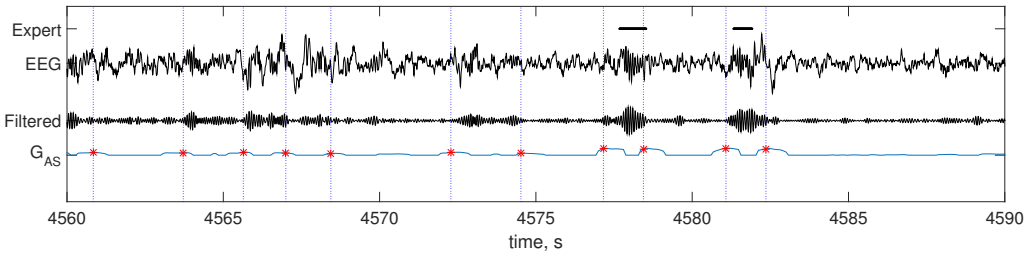


Figure 5.1: Adaptive segmentation process for based on filtered signal (Filtered) of original C3-M2 EEG (EEG) and corresponding expert scoring (Expert). G_{AS} represents segment detection function. Red asterisks represent local maxima of G_{AS}.

Table 5.1: Datasets details.

Dataset	number	fs (Hz)	time (m)	sleep spindles
Dreams V1	7	200/100	17.81 ± 3.08	42 ± 21
Dreams V2	5	200/100	16.53 ± 2.63	55 ± 23
NIMH	5	250	139.48 ± 29.39	182 ± 60

5.2.2 Proposed method: SSAD

The idea of proposed sleep spindle adaptive detector (SSAD) is to follow the definition of sleep spindle and adapt to the data at the same time. The method consists of two steps. In the first step it detects sleep spindle candidates, which fulfill the requirement to be an a 0.5–2-s burst in a sigma power. Then, we separate sleep spindles using a threshold obtained by GMM.

Adaptive segmentation is employed to obtain candidates to be a sleep spindle. The obtained candidate set is reduced by application of specific requirements. Adaptive segmentation is performed on EEG signal filtered in the sigma band. We state a sigma frequency band to be 11–16 Hz. Segmentation utilizes 0.5-s window and 0.1 of sampling frequency step. Std of the filtered signal is used as a signal metric function. Adaptive threshold is calculated as 0.5 of smoothed values obtained with moving average filter. Example of adaptive segmentation is shown in Figure 5.1. Then, we select all segments with length in $[0.3; 2]$ s to fulfill a sleep spindle length requirement. We use 0.3 s as a minimum duration to control error of adaptive segmentation ($\approx$ 2×step). In the next step, we fulfill condition that sleep spindle is a burst. To do this, we define a function M, which represents correlation with sigma power. We apply it to all segments and denote its outcome for a segment i as M_i. We select all segments s such $M_s > 0.25M_{s+1}$ and $M_s > 0.25M_{s-1}$ where $s-1$ and $s+1$ denote previous and subsequent to s segments

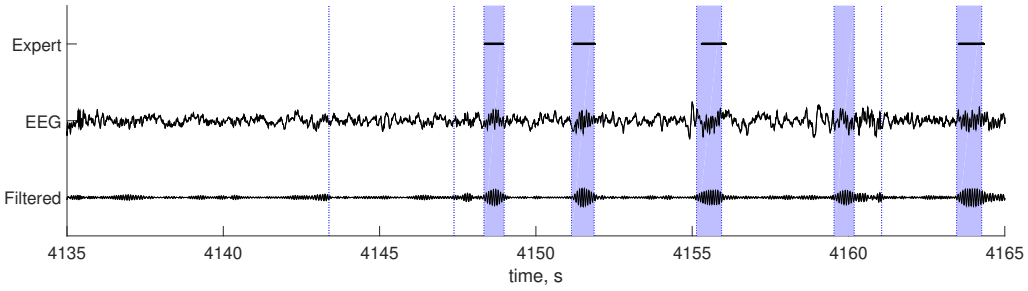


Figure 5.2: Sleep spindle candidates highlighted by blue color in C3-M2 EEG (EEG), filtered EEG and corresponding expert scoring (Expert). Vertical dashed lines represent adaptive segmentation result.

respectively. Signals with fine temporal resolution should use relative power in 11–16 Hz over 0.5–50 Hz as the function M. Segment std of filtered EEG is used otherwise. Example of candidates selected by this procedure is shown in Figure 5.2.

Set of sleep spindle candidates is selected. We assume that selected candidates belong to two groups: spindle and non-spindle. However, depending on the data, it is possible to use three clusters partition: spindle and two non-spindle clusters. That could be considered for more strict scoring or highly contaminated data. Therefore we calculate std of filtered signal within a segment and relative band bower in 11–16 Hz over 0.5–50 Hz to describe a candidate segment. We employ GMM to separate the candidate set into clusters. We select a cluster with the maximum amplitude to be a sleep spindle cluster. Example is shown in Figure 5.3.

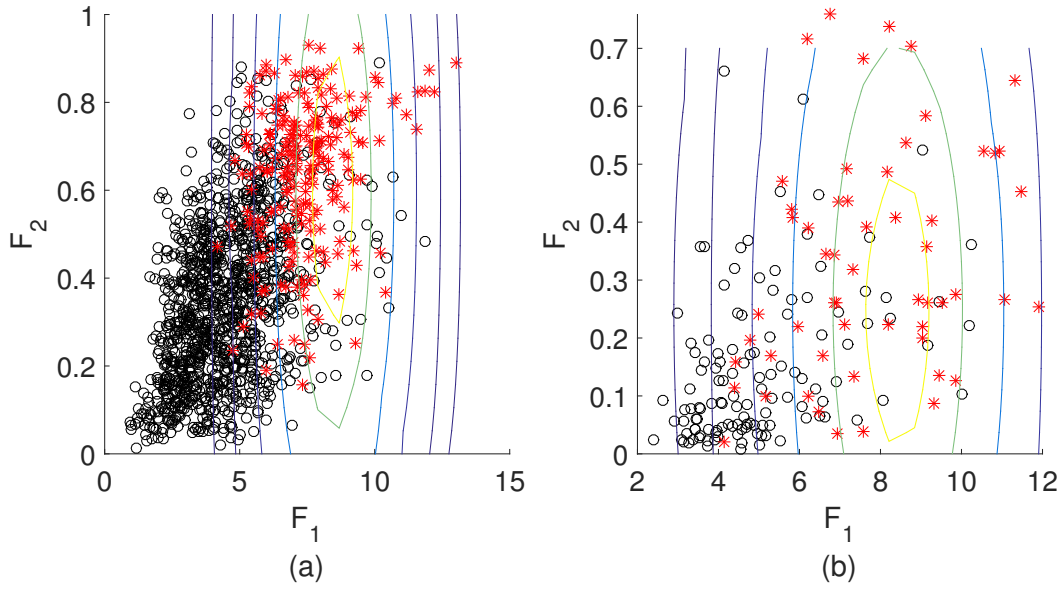
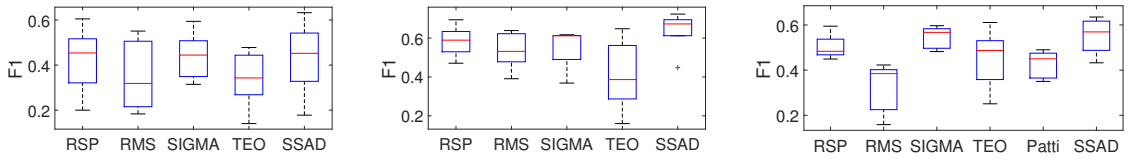


Figure 5.3: Feature distribution for NIMH (**a**) and Dreams V2 (**b**). F_1 represents std of filtered signal within a segment and F_2 is a relative sigma band power. Contours represent probability to belong to a sleep spindle cluster. Red asterisks represent spindle segments.



(a) (b) (c)

Figure 5.4: F_1 distribution for different methods obtained on Dreams V1 (**a**), Dreams V2 (**b**) and NIMH(**c**) datasets.

5.3 Experiment results

In this section we compare automatic sleep spindle detection methods. Performance evaluation utilize epoch based *Strategy 1* from Section 2.4.4 to discover ability of method to to detect events of interest. Recall *Rec*, precision *Prec* and measure $F1$ were calculated. Here, we use four sleep spindle detectors for testing purposes: RMS, RPS, SIGMA, TEO . There were used implementations of the methods by [95] for Python and [131] for MATLAB. The used threshold values were proposed in [95]. We do not solve problem of optimal threshold for NIMH dataset to emulate solving of a real world task. SSAD was applied to the data. It used division into 2 cluster for Dreams V1, V2 and 3 clusters NIMH dataset. Besides, we implemented the Patti method in MATLAB and used it in testing. The method required enough data to construct the sleep spindle cluster (in comparison to non-spindle cluster), therefore, testing results were obtained only for NIMH dataset. We added some modifications to reduce number of false positive detections. In particular, we removed segments where alpha index superiors the sigma index like it was in SIGMA. Additionally, we removed segments with sigma index lower than mean sigma index. The used window size was decreased to 0.5 s because using 1-s window reduces the difference between spindle and non-spindle sets. Finally, we repeatedly increased number of clusters until we obtained less than 500 spindles. Automatic sleep spindle detection by RMS, RPS, SIGMA, TEO, Patti and SSAD was applied to Dreams V1, V2 and NIMH datasets. The results are presented in Table 5.2, Table 5.3 and Table 5.4 respectively. Visual comparison of F_1 is provided in Figure 5.4.

The testing results in Dreams V1 show similar results except for TEO, which was too conservative and rejected too many events labeled by an expert. Lower *Rec* values reflect this fact. RMS also showed low average F_1. It otherwise returned more spindle event (97 ± 27) in comparison to expert labeling. Values obtained on Dreams V2 are much higher. TEO detection was too restricted again. It shows decreased *Rec* and, consequently, F_1. Similar to the previous dataset RMS detection returns too many events in comparison in expert labeling. However, increased *Rec* evidence about the better agreement with an expert. RSP and SIGMA in opposite are more conservative. However,

Table 5.2: Performance results for Dreams V1 dataset.

Method	F1	Rec	Prec
RMS	0.36 ± 0.15	0.67 ± 0.16	0.27 ± 0.10
RSP	0.43 ± 0.14	0.44 ± 0.18	0.45 ± 0.15
SIGMA	0.44 ± 0.10	0.47 ± 0.13	0.46 ± 0.15
TEO	0.34 ± 0.12	0.32 ± 0.17	0.53 ± 0.13
SSAD	0.43 ± 0.16	0.59 ± 0.15	0.37 ± 0.19

Table 5.3: Performance results for Dreams V2 dataset.

Method	F1	Rec	Prec
RMS	0.54 ± 0.10	0.75 ± 0.17	0.45 ± 0.13
RSP	0.58 ± 0.08	0.55 ± 0.16	0.67 ± 0.13
SIGMA	0.55 ± 0.11	0.49 ± 0.12	0.65 ± 0.17
TEO	0.41 ± 0.19	0.30 ± 0.18	0.83 ± 0.12
SSAD	0.64 ± 0.11	0.68 ± 0.11	0.63 ± 0.18

Table 5.4: Performance results for NIMH dataset.

Method	F1	Rec	Prec
RMS	0.31 ± 0.11	0.94 ± 0.07	0.2 ± 0.09
RSP	0.50 ± 0.06	0.47 ± 0.13	0.59 ± 0.13
SIGMA	0.54 ± 0.05	0.56 ± 0.11	0.55 ± 0.10
TEO	0.45 ± 0.14	0.80 ± 0.05	0.32 ± 0.13
Patti	0.43 ± 0.06	0.64 ± 0.13	0.33 ± 0.07
SSAD	0.55 ± 0.08	0.67 ± 0.13	0.52 ± 0.17

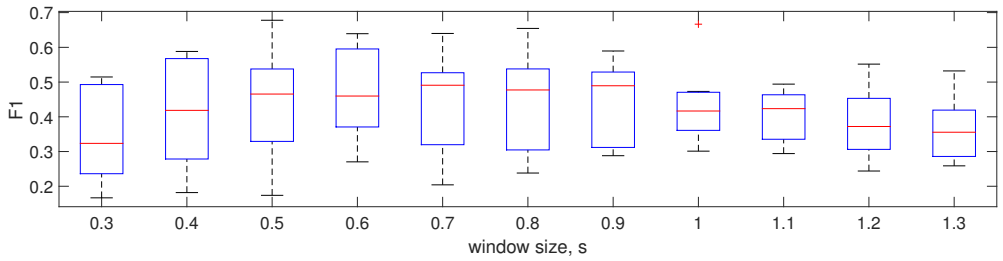
detection is more accurate. SSAD shows the best performance on this dataset in terms of F_1.

In the last NIMH dataset, RMS performed with the lowest F_1 across all methods. It shows the greatest *Rec* and extremely small *Prec*. Too many events were detected (884 ± 260). The chosen threshold was set too low. A similar tendency was observed in TEO. Methods RSP, SIGMA, and SSAD show similar results. However, obtained *Rec* values evidence RSP and SIGMA to detect more accurately among these methods, whereas SSAD detects more labeled events. Moreover, using GMM allows us to use probability estimation to belong to a spindle cluster. In the NIMH dataset, the expert acknowledged that the provided labels are a high-specificity/low-sensitivity evaluation of sleep spindles. Therefore, using the same cluster but with a lower radius will reduce the number of FP according to that expert labeling. Using optimal cluster radius increases F_1 to 0.64 ± 0.02.

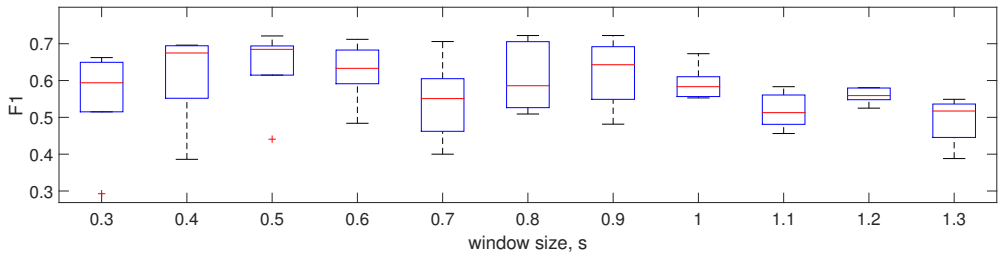
Performance of Patti is much lower than SSAD in terms of F_1. Mainly because it detects more events (299 ± 116), it is connected to the unequal clusters problem. Even after data reduction which we added, spindle cluster is still less than $2 \pm 3\%$ of the dataset used for clustering. For SSAD, this problem improved to $11 \pm 5\%$ due to candidate selection step. Moreover, clusters in Patti method overlap and false positive detections are controlled afterward by the duration requirements. Cluster separability in SSAD is much better. It is connected to advantages of adaptive segmentation which can hold signal property consistent within a segment, which increases class separability.

The greatest practical problem of adaptive segmentation is many parameters to tune. However, the window size is the most important one. Using small window size often results with a division of the event of interest into several smaller segments whereas using a large window size leads to missed events smaller than the window. In the testing, we used 0.5-s window because of the minimal sleep spindle duration. We performed testing used this assumption, but we did not know whether it is optimal. To investigate this problem, we iterate the window size in range 0.3–1.3 s with 0.1-s step and compute the F_1 to access the performance. The rest of the settings were the same. The obtained results are depicted in Figure 5.5. Assumption of the optimal window of size 0.5 s was confirmed for Dreams V2 dataset. In case of Dreams V1, any value in range 0.5–0.9 s was acceptable since the median value differed not more than by 0.02. Nevertheless, the optimal window choice in NIMH dataset is 0.7 s ($F_1 = 0.59 \pm 0.09$). That could be explained by the average spindle duration in this dataset (0.82 s) is signiicantly greater than 0.5.

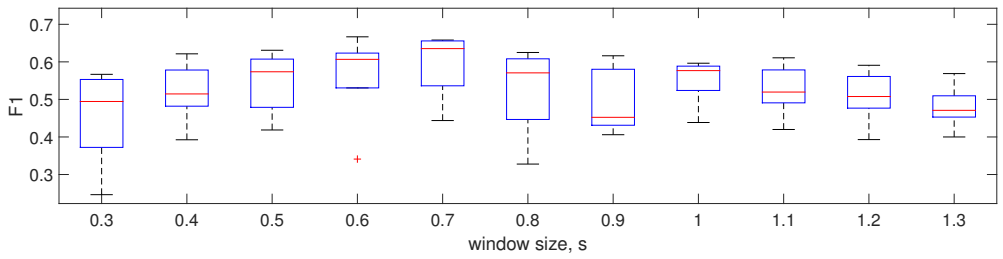
A Windows 10 Pro computer with an Intel Core i7 2.40-GHz CPU and 16 GB of RAM was used for testing. Sleep spindle detection for TEO, RMS, SIGMA and RSP took 7.3 ± 0.1 s, 9 ± 2 s, 59 ± 9.5 s and 73.7 ± 16.4 s respectively for NIMH dataset. The whole process in Patti method including computing of S-transform and modified clustering



(a)



(b)



(c)

Figure 5.5: F_1 distribution for different window size in SSAD obtained on Dreams V1 (**a**), Dreams V2 (**b**) and NIMH(**c**) datasets.

took 65 ± 12.08 s. And, finally, it was 8.8 ± 1.5 s for SSAD for the same dataset. The most expensive processes are power spectra calculation and S-transform. Execution time of GMM is acceptable for both SSAD (0.1 ± 0.05 s for clustering of 1542 ± 413 instances) and Patti methods (0.21 ± 0.01 s for clustering of 36895 ± 10733 instances).

5.4 Discussion

The adaptive sleep spindle detector was proposed and tested on real sleep EEG data. Obtained performance results were compared to the performance of previously developed methods. Statistic values for Dreams datasets are similar to published before [31], [95]. Using only sleep stage non-REM 2 could and different evaluation procedure could only slightly change the outcome. According to the expert labeling sleep spindles in NIMH dataset, the evaluation is very conservative. That corresponds to detecting only those events, in which the expert was sure. That fact may affect results obtained on NIMH

dataset. The problem of uncertain labeling suppresses the outcome results.

First methods of automatic sleep spindle detection were based on root mean square like RMS. Many studies still use this approach for spindle parameters estimation because many previous studies used it [41], [43], [88]. However, it is not robust to the noise since it analyzes only the sigma band. In case of muscular artifacts or arousals, the noise propagates to many high frequencies bands including sigma, which leads to false detections by RMS (and TEO as well). Unfortunately, it is not always possible to reject all the artifacts in clinical practice. That gives significant advantage for the methods investigating more comprehensive frequency range like SIGMA, RPS, Patti, and SSAD. However, strictly defined threshold like in RPS or SIGMA may restrict the outcome, especially in case of subjects suffering from different sleep disorders, which refers to both datasets. Therefore, the adaptive threshold can benefit in this case. There are two methods based on adaptive threshold both utilize the GMM model. The proposed SSAD outperforms the Patti method. Even they both utilize GMM, the way how they select data for clustering makes a difference. When Patti method cluster windows of the same length, the SSAD detects candidates, and it processes only chosen candidates. That trick allows the SSAD to decrease the number of elements. Moreover, selected segments have stationary properties within a segment due to the adaptive segmentation process. That improves separability in comparison to clustering of segments of equal length. Both fewer segments and better separability decreases the computational time and improves the results of SSAD in contrast to Patti method. Features extracted in Patti method also reflect "burst" requirement and spectral information, however, not optimal window length does not allow to achieve better performance. In the SSAD we combined all advantages of the previously proposed methods: using spectral information as a characteristic feature and adaptive threshold based obtained using GMM. Also, in the original paper, Patti method was tested and showed much better performance results on the dataset from other open-source database [94] where there were labeled more spindles per time unit. Unfortunately, we could not gain these data for our tests due to technical problems on the data provider side.

We believe that approach is very useful for sleep spindle detection in clinical work and detection of sleep EEG patterns in general. Defining a pattern sensitive function and employing it for adaptive segmentation will provide reasonable candidates for clustering. Using an adaptive threshold based on statistical distribution will adapt to the data and provide an all-embracing probabilistic framework for results evaluation. Changing the probability threshold we were able to adjust the results to be closer to the expert scoring.

5.4.1 Limitations and future work

One of the main limitations of the approach is the assumption on the data distribution. Therefore, using only part of the data or highly contaminated data may lead to a significant number of false results. Moreover, sleep spindles may occur in non-REM 3. However, we have no assumption on the distribution of candidates. This question requires more tests. That gives us one of the main directions for future work. Furthermore, we would like to investigate false positive results. They pass both steps of detection and are assumed as sigma bursts and having enough amplitude measure to be a sleep spindle. However, they were not labeled by the expert. Moreover, the agreement between different experts could be considered as moderate. One of the possible reasons, according to expert evaluating NIMH dataset is that sleep EEG data are too complex for visual analysis. We would like to investigate decomposition and visualization methods to make the manual scoring more reliable.

Manual scoring is a very time-consuming task, which limits the number of evaluated data recordings in the testing sets. In the future, we would like to test the proposed method on data from other public databases [94]. Moreover, we would like to evaluate the results in comparison to a combination of scorings provided by other experts for the same data. That may give more information on the agreement of experts and the best strategies for automatic evaluation. Finally, we would like to adapt the proposed method to events detection in other biomedical domains, especially EMG.

5.5 Conclusion

The study proposed a sleep spindle detector based on adaptive segmentation and adaptive threshold obtained using GMM. We fulfill the definition of sleep spindle by selecting candidates by the increase in spindle related information and length. Further division by amplitude features separates true sleep spindle class from the false events. The method is scalable and fully adaptive. The testing was performed on open-source datasets of short-term sleep EEG as well as on private long-term EEG. The testing results were compared with methods based on threshold and the previously proposed method using GMM for sleep spindle detection. Obtained results proved the effectiveness of the proposed method.

Chapter 6

Results Visualization for Practical Application

6.1 Introduction

This chapter is devoted to the practical problem of visualization of automatic pattern detection results. In the semiautomatic pipeline, an expert's task is to label patterns of interest using supporting information. In Chapter 3 an outcome of a chosen automatic detector was used for support of manual scoring. Using several detectors provides even more information for better decision making. However, simultaneous displaying several binary detection outcomes may be confusing for visual inspection, especially, in case of many detectors. Consulting with experts on how they perform scoring using supporting information reveals that their main interest is how many reliable detectors agreed on the decision. In this study, we investigate the problem of providing meaningful supporting information based on the outcomes of several detection methods. Moreover, we provide an approach for results visualization for practical applications and show its benefits for problems of EEG artifact and sleep spindle detection.

For the investigation, we first perform detection of the pattern of interest by several detectors. Then, we use voting for aggregation of results obtained from automatic EEG pattern detectors. It allows us to keep the architecture simple and, consequently, relevant for practical applications. Instead of a binary detection vector, we obtain a vector of values in the range [0,1], which represents the agreement of the detectors. However, voting utilizing equal weights does not reflect the "detector reliability" part of the expert's strategy. To fulfill this requirement, weights are assigned to detector outcomes in the voting procedure. Each weight corresponds to the accuracy of the detector. There are two possible approaches to perform this: supervised and unsupervised methods. Logistic regression is utilized to derive weights based on a supervised approach. The unsupervised

method comes from the field of crowd-sourcing, where there is suggested many raters with different accuracy are suggested. This approach is known for evaluation of human [101] and automatic [62] raters. It applies a generative Bayesian model for uncovering true hidden labels of the data and obtains the accuracy of each rater regarding uncovered true labels. Instead of using human raters as it was shown in [131] for sleep spindle detection problem, we use previously proposed automatic detectors to provide labeling of the data. In this chapter, we investigate the detection power of these three methods: unweighted one, weighted unsupervised, and weighted supervised approaches. Moreover, we show how we employ the selected methods in a real visualization application. The proposed approach could be used in other scientific fields. The publication is accepted for a conference: E. Saifutdinova, D. U. Dudysová, V. Gerla, and L. Lhotská, "Improvement sleep spindle detection by aggregation techniques", in *15th Mediterranean Conference on Medical and Biological Engineering and Computing (MEDICON2019)*, 2019.

6.2 Materials and methods

The task of pattern detection assumes that for a signal of N instances there are unknown real labels which can be represented as a binary vector $L = [l_1, \ldots, l_N]$. Each instance is established according to the task specificity and it could represent a signal sample, an epoch of a constant duration or a segment with variable length. L contains real pattern labels. Each value l_i is either 1 for an instance belonging to a searched pattern and 0 for a not relevant instance. In biomedical signal processing the vector L is often considered as unknown due to complexity of the field and the fact that obtained signals not always could be decoded. However, an expert scoring $S = [s_1, \ldots, s_N]$ could be used as an approximation of L. Values s_i are usually binary. Scoring obtained from independent experts could differ because of experience, trained skills and preferences for scoring of individual expert.

6.2.1 Unweighted aggregation

Aggregation approach assumes that X is a binary matrix of observations obtained from one recording of size $N \times M$ where N is the length of binary scoring and M is the number of detectors. Output of aggregation method is a binary detection vector $D = [d_1, \ldots, d_N]$. For a voting model decision d_i is based on the value v_i obtained as

$$v_i = \sum_{j=1}^{M} w_j X_{i,j} \tag{6.1}$$

where w_j is an impact of each j_{th} detector. Then,

$$d_i = \begin{cases} 0 & \text{if } v_i < T \\ 1 & \text{if } v_i \geq T \end{cases} \tag{6.2}$$

where T is a predefined threshold. The unweighted technique assigned equal weights w_i as $1/N$ to all detectors. That is the simplest approach to resolve an unknown weights problem. The unweighted model does not reflect performance differences in the methods which may not be optimal.

6.2.2 Unsupervised weighted aggregation

The goal of weighted model is to obtain a final decision which relies on importance of each detector. In this section we describe a fully unsupervised model. As in the previous studies [62], here we employ a Bayesian model to attempt to uncover the performance properties of the detectors in conditions of unknown real labels. We assume independence of each detector given binary vector of real labels L. Assuming that each j_{th} detector performs with TPR $\alpha_j \in [0, 1]$ and TNR $\beta_j \in [0, 1]$. Denote a row of X as x_i, it represents an outcome of M detectors for instance i. Then, x_i is dependent on a real label l_i and parameters of the detectors $\alpha = \{\alpha_1, \ldots, \alpha_M\}$ and $\beta = \{\beta_1, \ldots, \beta_M\}$. Also, we assume that values l_i come from Bernoulli distribution with unobserved parameter k. Following the [62], we regularize parameters k by a symmetric Beta distribution $Beta(\theta+1, \theta+1)$ for $\theta \geq 0$. All α_j and β_j are regularized by the same parameters φ and ψ using asymmetric Beta priors $Beta(1, \varphi + 1)$ and $Beta(1, \psi + 1)$.

Learning of the model is performed by a standard log-likelihood maximization strategy and EM method. Here, X is the set of observed values, L is a latent variable and k, α and β are the parameters. Knowing the parameters we can get [62]

$$\begin{aligned} p(L = 1|X; \alpha, \beta, k) &= \frac{p(X|L = 1; \alpha, \beta)p(L = 1; k)}{p(X|L = 1; \alpha, \beta)p(L = 1; k) + p(X|L = 0; \alpha, \beta)p(L = 0; k)} \\ &= [1 + \frac{1 - k}{k} \prod_j a_j^{X_j} b_j^{1-X_j}]^{-1}, \end{aligned} \tag{6.3}$$

where $a_j = \frac{\beta_j}{\alpha_j}$, $b_j = \frac{1-\beta_j}{1-\alpha_j}$ and X_j is the j_{th} column of X. Maximizing of $\ell(\alpha, \beta, k) = log\ p(X; \alpha, \beta, k)$ provides regularization of parameters [62] such as

$$k = \frac{\theta + \sum_i l_i}{2\theta + N} \tag{6.4}$$

$$\alpha_j = \frac{\sum_i l_i X_{i,j}}{\varphi + \sum_i l_i}. \tag{6.5}$$

$$\beta_j = \frac{\sum_i (1 - l_i) X_{i,j}}{\psi + \sum_i (1 - l_i)}. \tag{6.6}$$

Because L is a vector of Bernoulli variables, we update L by Equation 6.3 given parameters k, α_j and β_j [62]. The L is initialized as unweighted aggregation by majority voting. Steps are repeated until it converges. We assume convergence when $|k^\tau - k^{\tau+1}| + ||\alpha^\tau - \alpha^{\tau+1}|| + ||\beta^\tau - \beta^{\tau+1}|| < \epsilon$ where τ denotes iteration. We chose ϵ equal 10^{-12}.

6.2.3 Supervised weighted aggregation

The idea of supervised aggregation is based on learning a model on a set of provided examples. Due to high variability in sleep spindle density across all subjects, we suggest that a model should be trained on the data from the same patient. Following the strategy used in active learning [76], we suggest that an expert scores a signal from the beginning to some point on a time-line c. The signal before c is used for model training, and the obtained model is applied to the rest of the data. Then, testing divides the signal into two successive parts P_1 and P_2 by the cut point c and assumes that the expert labeling is provided only for P_1. After automatic detectors are applied to P_1 we obtain a binary matrix X' of size $N' \times M$ where N' is the number of instances in binary scoring of P_1 and M is the number of detectors. On the other hand, real labels for P_1 are known and could be interpreted as a binary vector Y' of length N'. Using X' and Y' we train a logistic regression with L_2 regularization and obtain an automatic scoring for the rest of the data by application of the obtained model to the P_2. Logistic regression was chosen due to its simplicity and a simple interpretation.

6.3 Experiment results

This section provides results of investigation of detection potential of all three methods. The results are presented and discussed for the tasks of sleep spindle detection and artifact detection.

6.3.1 Sleep spindle detection

First, we applied state-of-the-art sleep spindle detectors RMS, RSP, SIGMA, and TEO to datasets Dreams V1, Dreams V2, and NIMH. Obtained results are presented in the previous Chapter 5. The outcome of the detection is represented as a binary matrix. To

perform training of weighted voting models, we downsampled obtained binary detection using 0.5-s non-overlapped windows; labels were assigned by the majority of samples. All models were trained for separate subjects individually and do not depend on data of other subjects. The unsupervised model uses detection labels for the whole timeline, whereas supervised approach uses only the first part of continuous data. It contains 50% of sleep spindles for Dreams V1 and V2 datasets and 20% for NIMH dataset. Such values were selected based on signal length and consultancy with experts. Such division into training and testing data keeps the method relevant for practice in terms of data amount, which should be analyzed by an expert. Moreover, we excluded data with less than 25 sleep spindles from testing of the supervised weighed model due to the small number of training examples. Parameters for the unsupervised method were chosen empirically as ψ, φ equal 2 and θ is 1. Such parameters lead to faster convergence. Convergence was achieved not in more than 310 iterations for all datasets. Obtained weights further used for the combination of the outcome of separate detections. Results validation is performed by *Strategy 1* described in Section 2.4.4. Recall *Rec*, precision *Prec* and measure $F1$ were calculated. For all used approaches, $F1$, *Rec*, and *Prec* were calculated for the optimal threshold for each method. For weighted methods, we describe results for optimal threshold in terms of $F1$. Results for Dreams V1, Dreams V2 and NIMH datasets are presented in Tables 6.1, 6.2 and 6.3 respectively.

Table 6.1: Performance results for Dreams V1 dataset.

Method	F1	Rec	Prec
Unweighted	0.50 ± 0.09	0.61 ± 0.16	0.45 ± 0.14
Weighted unsupervised	0.50 ± 0.08	0.62 ± 0.15	0.46 ± 0.16
Weighted supervised	0.44 ± 0.10	0.52 ± 0.14	0.41 ± 0.17

Table 6.2: Performance results for Dreams V2 dataset.

Method	F1	Rec	Prec
Unweighted	0.65 ± 0.04	0.70 ± 0.14	0.64 ± 0.11
Weighted unsupervised	0.67 ± 0.04	0.69 ± 0.14	0.69 ± 0.14
Weighted supervised	0.68 ± 0.06	0.71 ± 0.13	0.67 ± 0.08

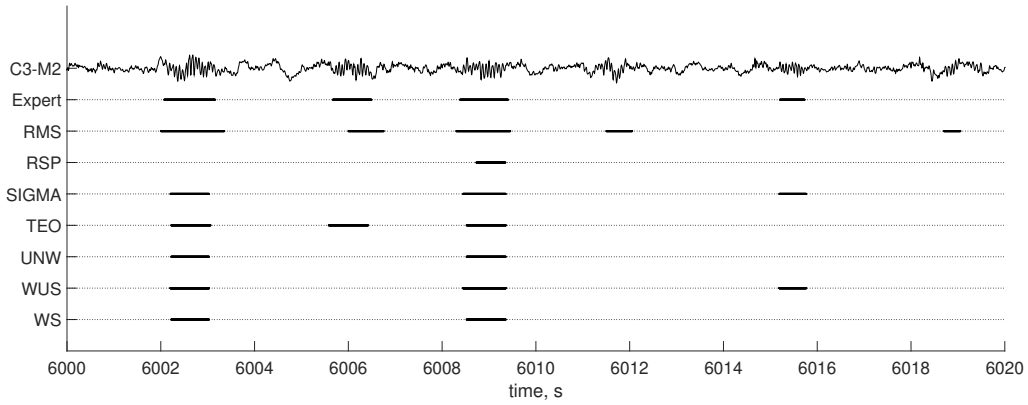


Figure 6.1: Example of sleep spindles in expert labeling (Expert) and detections performed by single detectors (RMS, RSP, SIGMA and TEO) and their combination using unweighted (UNW) and weighted unsupervised (WUS) and supervised (WS) approaches.

Table 6.3: Performance results for NIMH dataset.

Method	F1	Rec	Prec
Unweighted	0.56 ± 0.06	0.62 ± 0.14	0.54 ± 0.13
Weighted unsupervised	0.57 ± 0.05	0.62 ± 0.10	0.54 ± 0.09
Weighted supervised	0.56 ± 0.03	0.64 ± 0.21	0.54 ± 0.12

Methods of automatic sleep spindle detection and their combination were applied to three datasets. Combination of methods dramatically outperforms each sleep spindle detector in terms of $F1$ for all datasets. The unsupervised weighted method demonstrates slightly better results in comparison to unweighted unsupervised and supervised methods. The results of the weighted unsupervised method were reliable in terms of weights which were consistent with the previous results. It ended up with equally high (over 0.85 in average for each dataset) TPR and various TNR, which is reflected in the behavior of single detectors regarding false positive results. Moreover, corresponding TPR values for weighted unsupervised method are much greater than for other aggregation methods. Example of detection by all methods is presented in Figure 6.1. The unweighted method assumes a spindle label when at least two detectors agreed for Dreams V1 and V2 and agreement of three detectors for NIMH dataset. Weighted methods assign weights to the combination of detectors outcomes. Such as the weighted method in Figure 6.1 assigned greater weight to SIGMA detector and detects a sleep spindle at 6015 s. However, weight for sententious detection of RMS and TEO is low, and spindle at 6016 s was not discovered.

Many studies use one single method for discovering of sleep spindle [80], [88] due to the simplicity of the approach. However, detectors union can discover $68.7 \pm 30.8\%$ of the

sleep spindles for the Dreams V1, $80.2 \pm 5.2\%$ for Dreams V2 and $97.4 \pm 2.4\%$ for NIMH dataset. Using aggregation of detectors is a logical step to do. Results of aggregation methods sharply differ for all events if they are limited by the performance of separate detectors. However, the performance of aggregation methods is quite similar in terms of $F1$. The significant number of FP is a typical problem of all sleep spindle detectors and using $F1$ helps to balance both $Prec$ and Rec to obtain detection relevant for practice.

Uncertainty of expert evaluation of sleep spindles can limit performance the combination of the methods. In the case of NIMH dataset, the expert labeled only visible sleep spindles that did not overlap with other characteristic waves and had a clear shape. Despite on uncertainty of this kind, expert labeling is the optimal choice for data evaluation. Computational time is also a big concern. Testing was performed on a Windows 10 Pro computer with an Intel Core i7 2.40-GHz CPU and 16 GB of RAM using Python 2.7. It takes 5–7 s for unsupervised model learning and less than 0.01 s for a supervised approach for NIMH dataset. The aggregation time is acceptable in absolute terms; however, including processing time from several detection methods, it could be a noticeable time.

6.3.2 Artifact detection

Testing for artifact detection task is performed on the data presented in Chapter 3. Data described in Chapter 4 were not used for testing due to the short duration of the recordings. Therefore, each recording of the testing set is described as a binary vector provided by an expert and four binary vectors provided from detectors AFAST, FASST, FASTER, and SCADM. Results for these methods are described in Chapter 3. Results obtained by application of all three described strategies are evaluated by the *Strategy 2* explained in Section 2.4.4. Cohen's kappa K, sensitivity Se, and false discovery ratio FDR were calculated. We applied all three methods to the whole data and used the optimal threshold to investigate the best possible outcome. Parameters for the unsupervised method were chosen empirically as ψ and φ as 1 and θ is 0.5. Convergence was achieved not in more than 360 iterations. The supervised method utilized the first 10, 20, 30, 40, and 50% of the data for model learning. Obtained results did not significantly change until reaching 40% in training data. Results for 40 and 50% of the data in the training set were similar. Therefore, we provide results for a supervised model only for 10 and 40% of data. Obtained statistic values are presented in Table 6.4.

Table 6.4: Performance results for artifact dataset.

Method	K	Se	FDR
Unweighted	0.62 ± 0.10	0.54 ± 0.12	0.13 ± 0.10
Weighted unsupervised	0.71 ± 0.10	0.76 ± 0.12	0.26 ± 0.16
Weighted supervised (10%)	0.72 ± 0.10	0.77 ± 0.11	0.23 ± 0.16
Weighted supervised (40%)	0.74 ± 0.12	0.80 ± 0.14	0.21 ± 0.15

Weighted methods outperform the unweighted method. However, they show no significant improvement against AFAST. This method showed noticeably better performance against the other detectors on the used dataset. Both weighted methods assigned it bigger weights, which improves the result in comparison to the weighted approach. Weights obtained with unsupervised model repeat behavior described in previous tests. Unlikely to the test on sleep spindle dataset, here, the method ended up with almost equally low TNR but different TPR, which reflected the performance of a single detector. TPR for AFAST is 0.91 ± 0.06, for FASTER 0.61 ± 0.15, 0.35 ± 0.19 for FASST and 0.55 ± 0.23 for SCADM. Such values overestimate the real values, but the proportions are right. The computational time of the unsupervised method mainly depends on the number of iterations and data size. Testing was performed on a Windows 10 Pro computer with an Intel Core i7 2.40-GHz CPU and 16 GB of RAM using Python 2.7. The time varies from 5.8 to 46.1 s (for 160 and 298 iterations respectively). For supervised model learning, it took less than 0.01 s.

The performance of single methods differs over sleep stages, especially the sleep stage non-REM 3. Therefore, we decided to investigate the performance of the aggregation approach on the separate stages. The supervised method is not considered for separate sleep stages due to a rare practical scenario where it will be known at least 10% from each sleep stage. Moreover, it requires some negative examples, and 10% of each stage does not guarantee that the condition will be fulfilled. It is not guaranteed for the whole recording either but is often achieved in practice. Therefore, we concentrate on unweighted and weighted unsupervised methods for separate sleep stages. We used the same settings for the unsupervised method as we used for the testing on the whole data. This time, the average number of iterations was less than 250 in all tests. However, for one recording in stage non-REM 2 and one (different) recording in stage non-REM 3, it took over 1000 iterations to converge. That would be connected with that strategy of choosing a starting vector was not optimal for these data. Results of testing are presented in Table 6.5.

The results for stages Rem, non-REM 1 and 2 are similar to ones obtained for data without division into sleep stages. However, for several data recordings for non-REM 3

Table 6.5: Performance results for artifact dataset for sleep stages.

Method	K	Se	FDR
Rem			
Unweighted	0.66 ± 0.17	0.59 ± 0.19	0.09 ± 0.12
Weighted unsupervised	0.74 ± 0.17	0.77 ± 0.20	0.18 ± 0.15
non-REM 1			
Unweighted	0.62 ± 0.11	0.57 ± 0.12	0.10 ± 0.08
Weighted unsupervised	0.74 ± 0.10	0.77 ± 0.13	0.15 ± 0.08
non-REM 2			
Unweighted	0.63 ± 0.13	0.58 ± 0.17	0.15 ± 0.18
Weighted unsupervised	0.75 ± 0.13	0.84 ± 0.11	0.24 ± 0.21
non-REM 3			
Unweighted	0.53 ± 0.25	0.46 ± 0.25	0.15 ± 0.16
Weighted unsupervised	0.48 ± 0.31	0.71 ± 0.27	0.38 ± 0.32

stages the unsupervised weighted method converged to weights opposite than expected. It had extremely low TPR and high TNR for all detectors. That decreased the mean value of the data. Non-REM 3 sleep stage typically has a minimal number of artifacts. Therefore, automatic detectors tend to label fewer artifacts. Lack of automatic detections in all methods could lead to such behavior.

6.4 Visualization application

In this chapter, we provide an overview of a visual application, which could be used in practical work. The aggregation methodology provides a vector $D = [d_1, \ldots, d_N]$ for a data $\in R^{M \times N}$. Each value in d_i is in range [0,1]. The d_i's could be interpreted as a probability to be a pattern of interest. It could be visually displayed using the color transparency as it is showed in Figure 6.2 for sleep spindle detection task.

However, it may take more efforts to embed this idea in an existing application. Further, we demonstrate examples of such visualization for the pipeline explained in Chapter 3. First, we have to extend the pipeline as it is shown in Figure 6.3. Several automatic detectors are applied to the data. Then, the binary results are automatically combined into one vector where each value could be interpreted as a probability to be an artifact. Here, we used unsupervised method due to its greater detection power and no training data requirement. We combined detectors without division on sleep stages to avoid the expert confusion and problems of lack of detections like it was for sleep stage non-REM 3. The visualization part is implemented using functions EEG Lab in MATLAB; the combination part was performed as an external module in Python.

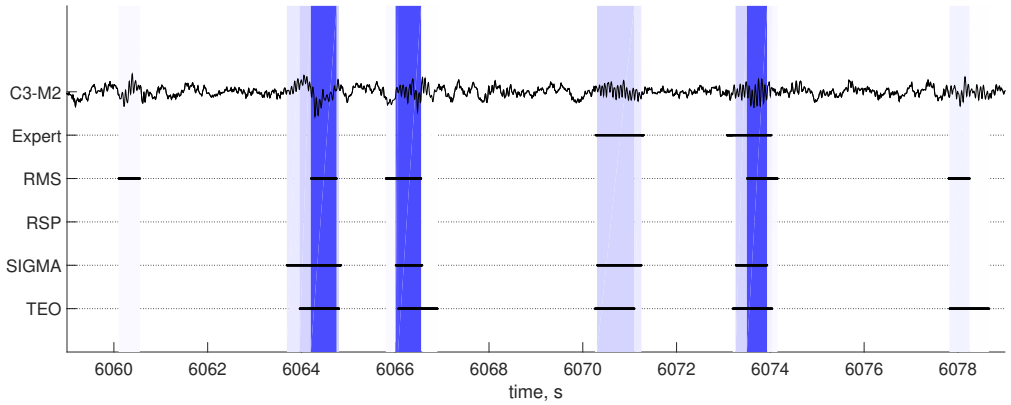


Figure 6.2: Example of sleep spindles in C3-M2 channel (C3-M2) labeled by an expert (Expert), detections by sleep spindle detectors (RMS, RPS, SIGMA, TEO) and reliability combination of detection by unsupervised method expressed intensity of the blue color (more intense color represent more reliable result).

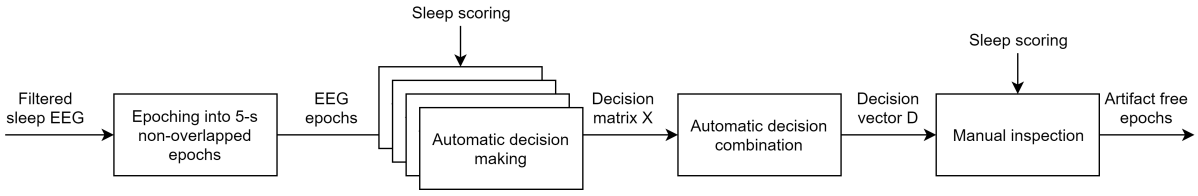


Figure 6.3: Extended pipeline for sleep EEG processing from Chapter 3.

For visualization part the programming code should be changed as it shown below.

```
command = ['nep = g.frames/g.trialstag; nsam = g.trialstag;'...
'[trialrej elecrej] = eegplot2trial(TMPREJ, nsam,'...
'nep,[',num2str(expertcolor),']);'...
'reject = trialrej; save(''reject.mat'',''reject'');'];

eegplot(data, 'command', command, 'wincolor', expertcolor,
 'srate', fs, 'spacing', 100, 'winlength', 18, 'winrej', events)
```

First, function *eegplot* should be used instead of *pop_eegplot*. As a first argument, it takes a data matrix $\in R^{M\times K\times N}$ where M is a channel number, K is a length of one epoch and N is the number of epochs as the input parameter. Data dimensions could be derived using internal variable of the function (g) as it is shown. The sampling frequency is determined by variable *fs*. Then, events for printing are defined using the variable *events*, which defines a matrix, and each row describes an event. It includes an event start and end in samples, three positions for color and M zeros required for highlighting a channel (not required, therefore, zeros). The expert color is significantly different. The colors provide only supporting information. Therefore, the variable *expertcolor* plays a huge role in decoding labeled events. It is an array of length 3, and it represents a color, which marks manually chosen epochs. We add it in the parameters of *eegplot* and the

Figure 6.4: Color scheme used in the visual application.

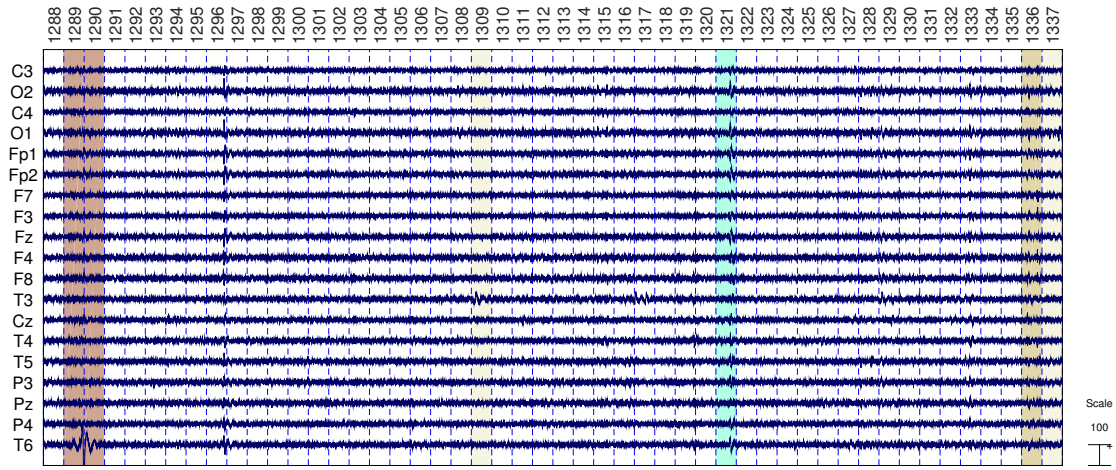
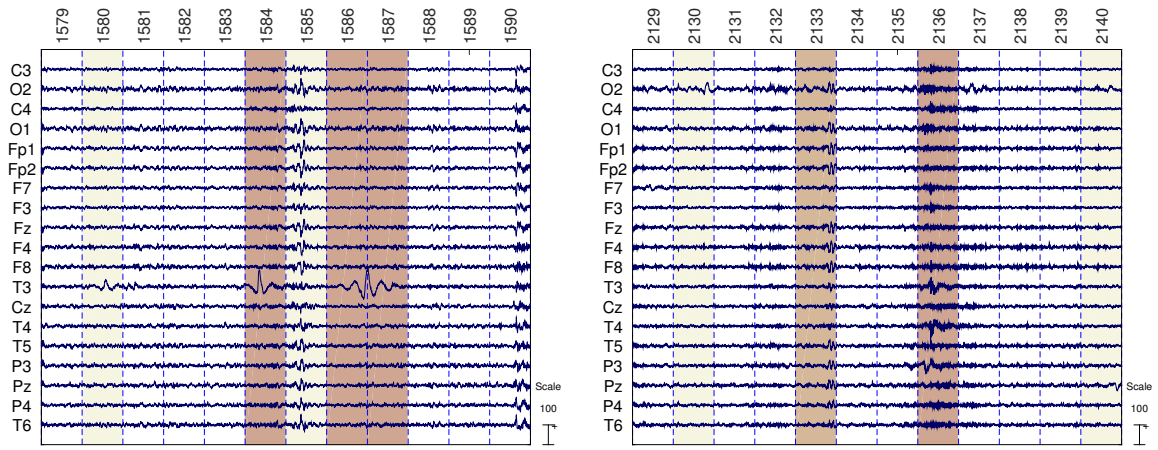


Figure 6.5: Reliability of the detection (using unsupervised weighted aggregation) visualization for artifact detection. Darker color represents a higher probability that it is an artifact. Light blue color (epoch 1321) is used for expert scoring. Vertical dashed lines separate 5-s epochs, epoch numbers are given above.

command variable as a *eegplot2trial* parameter. That parameter indicates, which events will be stored into *reject.mat* file. Colours are extremely important for visualization. We used a standard color of EEG Lab for epoch selection and the second half of colormap “pink" (for lighter colours) for representing combination results (shown in Figure 6.4). Figure 6.5 displays an example with automatic labeling and user selection. Figure 6.6 displays more detailed examples.

6.5 Discussion

A visualization methodology is presented in the study. We follow the expert strategy of agreement of reliable detectors. We investigate three approaches based on aggregation to satisfy this strategy on sleep spindle and artifact datasets. The unweighted approach proved to be the easiest for understanding and implementation. It has a great advantage when the performance of single detectors is equal. However, weighted methods have strong benefits over the unweighted method. Weights could be interpreted as a probability to be a pattern of interest. That gives more insight into the data and detection during visual analysis. The main benefit of the unsupervised method is that they do



(a) (b)

Figure 6.6: Reliability (by unsupervised weighted aggregation) visualization for artifact detection for channel loss of contact (**a**) and high frequency artifacts (**b**). Darker colors are assigned by color scheme in Figure 6.4. Vertical dashed lines separate 5-s epochs, epoch numbers are given above.

not need any information for model trainig. It attempts to uncover hidden labels L. Such a task is valuable for any EEG pattern detection task. However, having detectors with low performance, it is really hard to achieve very good results as it was shown for sleep spindle detection task. The supervised method tends to outperform unsupervised methods. Performance of the methods is directly connected to expert labeling. While unsupervised method tries to detect the real labels, the supervised method adapts to the expert decision, which may increase performance. However, the lack of training examples may negatively affect results.

Weighted methods are very powerful and may be used for a system for visual investigation. Despite on benefits of the supervised approach is less practical than the unsupervised approach due to the requirement of data for model learning. However, such an approach still could be used for weights correction. The unsupervised method was embedded in the practical application for artifact detection. We build this methodology into the pipeline presented in the previous Chapter 3. Displayed probability serves as supporting information for visual inspection. We do not solve the problem of optimal threshold since there is provided a visual interpretation of the colors. Moreover, the unsupervised method attempts to predict the real labels of the data, which may give the expert more trust to the application.

6.5.1 Limitations and future work

Despite all the benefits provided by the unsupervised method, there are still open questions. That gives us directions for future work. First of all, we would like to investigate in details effect of regularization scheme on performance. The original publication [62] used additional parameters which were not included in our tests due to the simplicity of the solution. However, in some tests, it assigned extremely low values for TNR and TPR, which could be avoided by adding clipping parameters in a convergence condition. Introducing a new regularization scheme may decrease computational time or improve performance. Testing in artifact detection task in non-REM 3 showed that the insufficient number of detected events might lead to unexpected results. Therefore, we would like to concentrate on the investigation of conditions where the solution converges. Especially in terms of what should be an agreement between detectors and how many labels should be uncovered for a stable result.

Future work will be concentrated on the further development of the application presented in this chapter. One of the main questions is how it affects expert scoring. Investigation, in which cases the expert trust the application more, will allow us to model the expert's behavior better and, consequently, provide better supporting information. Moreover, we investigate the connection between experts agreement and the aggregation performance. In testing, there are presented performance of sleep spindle detectors aggregation regarding two experts. More sophisticated techniques dealing with this problem of spindle detection evaluation like in [31], [73] will be employed. It will provide more accurate labeling of events labeled as sleep spindle and benefit on weights obtaining step. Furthermore, we include more detectors to improve detection results of aggregation. Also, we will test the effect of including several detectors with different parameters.

6.6 Conclusion

The chapter describes the combination of multiple detectors outcome for supporting visual inspection performed by an expert. Following the expert strategy, we rely on aggregation of several detectors. We test unweighted, weighted supervised, and unsupervised approaches. Advantages, disadvantages, and practical benefits of all presented methods are discussed. Testing is performed for data from sleep spindle detection and artifact detection tasks. Visualization application for practical use is presented.

Chapter 7

Summary and Perspective

In this study, we investigated an automatic sleep EEG pattern detection. We concentrated on data with high inter-subject variability, Specifically, on data collected from subjects suffering from insomnia and other sleep disorders. Variety of normal patterns and artifacts in sleep EEG is determined by the type and severity of the sleep disorder. All the data were collected on the specialized clinics and directly related to the real practice. In the study, we focused on artifact rejection and sleep spindle detection problems. Performing testing and achieving stable results for such data makes a great impact on understanding the issues in the field and providing better solutions. Using recent advances in pattern detection in EEG and other fields, we proposed reliable applications to help research in the real sleep laboratory. In our applications, we balanced the complexity of the solution and practical relevance. We used expert knowledge and practical experience to make it possible. Besides the theoretical description, we provide some details on the implementation, which allows the reader to repeat the experiments. Data collected in the National Institute of Mental Health, Czech Republic, cannot be published; however, the most of experiments could be repeated on data from open-source Dreams database.

7.1 Thesis achievement

We will now report on the extent to which the research aims set in Section 1.1 have been accomplished; they can be summarized as follows:

- The problem of automated pattern detection in sleep EEG recorded as a part of whole night PSG was investigated. A processing pipeline for sleep PSG/EEG data processing and pattern detection in a sleep laboratory was described in Chapter 3. State-of-the-art methods were tested on data collected from patients suffering from sleep disorders. The obtained results were described in Chapters 3, 4 and 5. Moreover, we investigate an expert's strategy for manual data inspection. We based an

application described in Chapter 6 to follow this strategy.

- The problem of artifact detection in a long-term EEG for a practical application has been studied. An extension of the previously published state-of-the-art method for artifact detection was proposed. The method utilizes clustering, which provides a more adaptive interface and makes it suitable for data with high inter-subject variability. The method was tested on a private dataset collected from subjects suffering from insomnia and compared to results of previously developed methods. Method details and testing results are presented in Chapter 3.

- Spatial properties of normal sleep EEG patterns and artifacts were studies. A novel automatic abnormality detection method using Riemannian geometry was proposed. The proposed method was tested on real sleep EEG from subjects suffering from sleep disorders, including data obtained from the open-source Dreams database and 19-channel sleep EEG private dataset recorded in a sleep research laboratory. The method details and experiment results are described in Chapter 4. Moreover, the developed toolbox is available online at `https://gitlab.ciirc.cvut.cz/open-source/rps`.

- The problem of adaptive segmentation for automatic sleep spindle detection was studied. The proposed method is based on the GMM clustering model for adaptive threshold identification. Both adaptive segmentation and adaptive threshold by GMM increase the ability of the method to adapt to the data. Testing was performed on real EEG from subjects suffering from insomnia, including data obtained from the open-source Dreams database and non-REM 2 sleep EEG from a private dataset obtained in a sleep research laboratory. Results were compared with the results obtained using previously developed methods, including another approach based on GMM. The method details and experimental results are described in Chapter 5.

- The problem of aggregation of results obtained from automatic EEG pattern detectors was studied for the sleep spindle detection and artifact rejection problems. The used aggregation method, namely voting allowed us to keep the method simple to be relevant for practice. We investigated the detection power of three different approaches for voting, including the unweighted model, weighted unsupervised, and weighted supervised models. Testing was performed on datasets presented in the previous chapters. Advantages and disadvantages of this approach were discussed in Chapter 6. Moreover, an improvement for practical application for visual data scoring using supporting information was proposed and embedded into the processing pipeline described in Chapter 3.

7.2 Future work

In future work, we would like to concentrate on the following problems. First, we are going to continue working on the presented methods, especially the ones, that were used for applications described in Chapters 3 and 6. Some directions for future work was mentioned in the previous chapters. We are planning to include the proposed methods into upcoming sleep research projects in the sleep research laboratory in the National Institute of Mental Heath, Czech Republic. In perspective, it will require further improvements and adjustments of the methods. We are going to test the adaptive ability of the methods in future studies, which include analysis of data collected from healthy controls as well as from subjects suffering from parasomnia and insomnia. We would like to collect representative groups of good sleepers and subjects who have sleep disorders for investigation of the difference in EEG patterns. Numerous medical studies are devoted to this problem; however, they do not provide overview from the technical point of view.

Another interesting aspect for a future study is extending of the methods to non-EEG PSG data. This is one of the most promising directions for future work. The method proposed for sleep spindle detection in EEG could be adapted to other domains such as tonic and phasic activity detection in EMG. That is an important EMG pattern which plays a big role in the investigation of RBD and other sleep-related movement disorders. Only a few detectors were proposed for detection of this patterns [58], [84]. The adaptation of the method presented in Chapter 5 is currently under investigation for this task. Our preliminary case study experiments have shown promising results. Our current work is concentrated on data collecting and scoring. Further, we would like to extend this method using context analysis to breathing artifact detection/rejection and PLMs detection in RBD patients.

In future, we would like to concentrate on studying of spatial patterns in parasomnia. The sleep stages mixturing is typical for this disorder. However, previous studies [71] confirmed the effectiveness of using spatial patterns for the sleep stage scoring. Our idea is to extend the method described in Chapter 4 for detection of stage mixtures. EEG patterns clusters corresponding to normal activity are constructed by performing artifact detection. Comparing data to such clusters allows for estimation of the probability to be a specific sleep stage. We expect that for sleep stage dissociation this probability will rapidly oscillate between two stages. We also are going to extend the method to other domains, particularly EMG and EOG. We believe that research on spatial patterns in non-EEG domains and their connection to events in EEG data will improve the detection accuracy.

Finally, we would like to continue working on practical applications. Our applications are based on the free open-source EEG Lab software with a user-friendly graphical inter-

face which is very popular in EEG research laboratories across the world. However, most of the functions of EEG Lab are devoted to processing of wake short-term EEG. EEG Lab is powerful enough to make the sleep data processing possible. However, it is not always obvious how to employ internal structure and functions for sleep research properties, especially for beginners and non-technical specialists. We would like to publish our practices for sleep data processing as a set of step-by-step tutorials. We also would like to make a plugin for EEG Lab, which allows to create EEG Lab data structures and make some processing steps automatically.

Bibliography

[1] U. R. Acharya, F. Molinari, S. V. Sree, S. Chattopadhyay, K.-H. Ng, and J. S. Suri, "Automated diagnosis of epileptic eeg using entropies", *Biomedical Signal Processing and Control*, vol. 7, no. 4, pp. 401 –408, 2012, ISSN: 1746-8094.

[2] A. Acharyya, P. N. Jadhav, V. Bono, K. Maharatna, and G. R. Naik, "Low-complexity hardware design methodology for reliable and automated removal of ocular and muscular artifact from eeg", *Computer Methods and Programs in Biomedicine*, vol. 158, pp. 123 –133, 2018, ISSN: 0169-2607. DOI: 10.1016/j.cmpb.2018.02.009.

[3] N. Acir and C. Guzelis, "Automatic recognition of sleep spindles in eeg by using artificial neural networks", *Expert Systems with Applications*, vol. 27, no. 3, pp. 451–458, 2004.

[4] C. C. Aggarwal, *Data mining: The textbook*. Springer Publishing Company, Incorporated, 2015, ISBN: 3319141414, 9783319141411.

[5] B. Ahmed, A. Redissi, and R. Tafreshi, "An automatic sleep spindle detector based on wavelets and the teager energy operator", in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Sep. 2009, pp. 2596–2599. DOI: 10.1109/IEMBS.2009.5335331.

[6] B. Ahmed, A. Redissi, and R. Tafreshi, "An automatic sleep spindle detector based on wavelets and the teager energy operator", in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Sep. 2009, pp. 2596–2599.

[7] F. Alimardani, R. Boostani, and M. Taghavi, "Classification of bmd and schizophrenic patients using geometrical analysis of their eeg signal covariance matrices", in *2015 38th International Conference on Telecommunications and Signal Processing (TSP)*, Jul. 2015, pp. 1–5. DOI: 10.1109/TSP.2015.7296459.

[8] M. N. Anastasiadou, M. Christodoulakis, E. S. Papathanasiou, S. S. Papacostas, and G. D. Mitsis, "Unsupervised detection and removal of muscle artifacts from scalp eeg recordings using canonical correlation analysis, wavelets and random forests", *Clinical Neurophysiology*, vol. 128, no. 9, pp. 1755 –1769, 2017, ISSN: 1388-2457. DOI: 10.1016/j.clinph.2017.06.247.

[9] A. Andreev, A. Barachant, F. Lotte, and M. Congedo, "Recreational applications of openvibe: Brain invaders and use-the-force", *BrainComputer Interfaces 2: Technology and Applications*, 241–257, 2016.

[10] S. M. Anisheh and H Hassanpour, "Adaptive segmentation with optimal window length scheme using fractal dimension and wavelet transform", *JE TRANSACTIONS B: Applications*, vol. 22, no. 3, pp. 257–268, Oct. 2009.

[11] I. Arnulf and C. Vernet, "Idiopathic Hypersomnia with and without Long Sleep Time: A Controlled Series of 75 Patients", *Sleep*, vol. 32, no. 6, pp. 753–759, Jun. 2009, ISSN: 0161-8105. DOI: 10.1093/sleep/32.6.753.

[12] A. Barachant, A. Andreev, and M. Congedo, "The riemannian potato : An automatic and adaptive artifact detection method for online experiments using riemannian geometry", 2013.

[13] Q. Barthélemy, L. Mayaud, D. Ojeda, and M. Congedo, "The riemannian potato field: A tool for online signal quality index of eeg", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 2, pp. 244–255, Feb. 2019, ISSN: 1534-4320. DOI: 10.1109/TNSRE.2019.2893113.

[14] M. H. Bonnet, "Effect of Sleep Disruption on Sleep, Performance, and Mood", *Sleep*, vol. 8, no. 1, pp. 11–19, Mar. 1985, ISSN: 0161-8105. DOI: 10.1093/sleep/8.1.11.

[15] V. Bono, S. Das, W. Jamal, and K. Maharatna, "Hybrid wavelet and emd/ica approach for artifact suppression in pervasive eeg", *Journal of Neuroscience Methods*, vol. 267, pp. 89 –107, 2016, ISSN: 0165-0270. DOI: 10.1016/j.jneumeth.2016.04.006.

[16] J. Born, B. Rasch, and S. Gais, "Sleep to remember", *The Neuroscientist*, vol. 12, no. 5, pp. 410–424, 2006, PMID: 16957003. DOI: 10.1177/1073858406292647.

[17] A. Borowicz, "Using a multichannel wiener filter to remove eye-blink artifacts from eeg data", *Biomedical Signal Processing and Control*, vol. 45, pp. 246 –255, 2018, ISSN: 1746-8094. DOI: 10.1016/j.bspc.2018.05.012.

[18] R. Cespuglio, D. Colas, and S. Gautier-Sauvigné, "Energy processes underlying the sleep–wake cycle", in *The Physiologic Nature of Sleep*, pp. 3–21. DOI: 10.1142/9781860947186_0001.

[19] C. Chang, S. Hsu, L. Pion-Tonachini, and T. Jung, "Evaluation of artifact subspace reconstruction for automatic eeg artifact removal", in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul. 2018, pp. 1242–1245. DOI: 10.1109/EMBC.2018.8512547.

[20] A. Chatburn, S. Coussens, D. Kennedy, K. Lushington, M. Kohler, and M. Baumert, "Sleep Spindle Activity and Cognitive Performance in Healthy Children", *Sleep*, vol. 36, no. 2, pp. 237–243, Feb. 2013, ISSN: 0161-8105. DOI: 10.5665/sleep.2380.

[21] X. Chen, Q. Chen, Y. Zhang, and Z. J. Wang, "A novel eemd-cca approach to removing muscle artifacts for pervasive eeg", *IEEE Sensors Journal*, pp. 1–1, 2018, ISSN: 1530-437X. DOI: 10.1109/JSEN.2018.2872623.

[22] X. Chen, H. Peng, F. Yu, and K. Wang, "Independent vector analysis applied to remove muscle artifacts in eeg data", *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 7, pp. 1770–1779, Jul. 2017, ISSN: 0018-9456. DOI: 10.1109/TIM.2016.2608479.

[23] X. Chen, X. Xu, A. Liu, M. J. McKeown, and Z. J. Wang, "The use of multivariate emd and cca for denoising muscle artifacts from few-channel eeg recordings", *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 2, pp. 359–370, Feb. 2018, ISSN: 0018-9456. DOI: 10.1109/TIM.2017.2759398.

[24] X. Chen, A. Liu, H. Peng, and R. K. Ward, "A preliminary study of muscular artifact cancellation in single-channel eeg", *Sensors*, vol. 14, no. 10, pp. 18 370–18 389, 2014, ISSN: 1424-8220. DOI: 10.3390/s141018370.

[25] H. R. Colten, C. o.S. M. Bruce M. Altevogt Editors, and Research, *Sleep disorders and sleep deprivation: An unmet public health problem*. National Academies Press (US), 2006, ISBN: 0-309-65727-X.

[26] M. Congedo, R. Phlypo, and A. Barachant, "A fixed-point algorithm for estimating power means of positive definite matrices", in *2016 24th European Signal Processing Conference (EUSIPCO)*, Aug. 2016, pp. 2106–2110. DOI: 10.1109/EUSIPCO.2016.7760620.

[27] M. Congedo, P. L. C. Rodrigues, F. Bouchard, A. Barachant, and C. Jutten, "A closed-form unsupervised geometry-aware dimensionality reduction method in the riemannian manifold of spd matrices", in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul. 2017, pp. 3198–3201. DOI: 10.1109/EMBC.2017.8037537.

[28] M. Congedo, "Eeg source analysis", Habilitation thesis, Doctoral School EDISCE, University of Grenoble., 2013.

[29] M. Congedo, A. Barachant, and R. Bhatia, "Riemannian geometry for eeg-based brain-computer interfaces; a primer and a review", *Brain-Computer Interfaces*, vol. 4, no. 3, pp. 155–174, 2017. DOI: 10.1080/2326263X.2017.1297192.

[30] A. Delorme and S. Makeig, "Eeglab: An open source toolbox for analysis of single-trial eeg dynamics including independent component analysis", *Journal of Neuroscience Methods*, vol. 134, pp. 9–21, 2004.

[31] S. Devuyst, T. Dutoit, P. Stenuit, and M. Kerkhofs, "Automatic sleep spindles detection — overview and development of a standard proposal assessment method", in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2011, pp. 1713–1716. DOI: 10.1109/IEMBS.2011.6090491.

[32] S. Devuyst, T. Dutoit, T. Ravet, P. Stenuit, M. Kerkhofs, and E. Stanus, "Automatic processing of eeg-eog-emg artifacts in sleep stage classification", 2008.

[33] K. Dhindsa, "Filter-bank artifact rejection: High performance real-time single-channel artifact detection for eeg", *Biomedical Signal Processing and Control*, vol. 38, pp. 224 –235, 2017, ISSN: 1746-8094. DOI: 10.1016/j.bspc.2017.06.012.

[34] S. P. Drummond and G. G. Brown, "The effects of total sleep deprivation on cerebral responses to cognitive performance", *Neuropsychopharmacology*, vol. 25, no. 5, Supplement 1, S68 –S73, 2001, Recent Advances in Sleep and Chronobiology, ISSN: 0893-133X. DOI: 10.1016/S0893-133X(01)00325-6.

[35] D. Dudysova, K. Veldova, M. Smotek, E. Saifutdinova, J. Koprivova, J. Buskova, B. A. Mander, M. Brunovsky, P. Zach, J. Korcak, V. Andrashko, M. Viktorinova, F. Tyls, A. Bravermanova, T. Froese, T. Palenicek, and J. Horacek, "Effects of daytime administration of psilocybin on sleep: Similar changes to antidepressants?", *Frontiers in Pharmacology*, 2019.

[36] D. Dudysová, I. Malá, K. Mladá, E. Saifutdinova, J. Koprivova, and P. Sos, "Structural and construct validity of the czech version of the pittsburgh sleep quality index in chronic insomnia", *Neuro endocrinology letters*, vol. 38, no. 1, 67—73, Feb. 2017, ISSN: 0172-780X.

[37] F. Duman, A. Erdamar, O. Erogul, Z. Telatar, and S. Yetkin, "Efficient sleep spindle detection algorithm with decision tree", *Expert Systems with Applications*, vol. 36, no. 6, pp. 9980 –9985, 2009, ISSN: 0957-4174. DOI: 10.1016/j.eswa.2009.01.061.

[38] R Esteller, "A comparison of waveform fractal dimension algorithms", *Circuits and Systems I: Fundamental Theory and Applications*, vol. 48, no. 2, pp. 177–183, Feb. 2001.

[39] M. Fantini, M. Michaud, N. Gosselin, G. Lavigne, and J. Montplaisir, "Periodic leg movements in rem sleep behavior disorder and related autonomic and eeg activation", *Neurology*, vol. 59, no. 12, pp. 1889–1894, 2002, ISSN: 0028-3878. DOI: 10.1212/01.WNL.0000038348.94399.F6.

[40] F Ferrarelli and et al., "Reduced sleep spindle activity in schizophrenia patients", *The American Journal of Psychiatry*, vol. 164, no. 3, pp. 483–92, Mar. 2007.

[41] F. Ferrarelli, R. Huber, M. J. Peterson, M. Massimini, M. Murphy, B. A. Riedner, A. Watson, P. Bria, and G. Tononi, "Reduced sleep spindle activity in schizophrenia patients", *American Journal of Psychiatry*, vol. 164, no. 3, pp. 483–492, 2007, PMID: 17329474. DOI: 10.1176/ajp.2007.164.3.483.

[42] C Fournier and D Inkpen, "Segmentation similarity and agreement", in *Proc. of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Stroudsburg, PA, 2012, pp. 152–61.

[43] S. Gais, M. Mölle, K. Helms, and J. Born, "Learning-dependent increases in sleep spindle density", *Journal of Neuroscience*, vol. 22, no. 15, pp. 6830–6834, 2002, ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.22-15-06830.2002.

[44] J. Gan and Y. Tao, "Dbscan revisited: Mis-claim, un-fixability, and approximation", in *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Victoria, Australia*, ser. SIGMOD '15, ACM, 2015, pp. 519–530, ISBN: 978-1-4503-2758-9. DOI: 10.1145/2723372.2737792. [Online]. Available: http://doi.acm.org/10.1145/2723372.2737792.

[45] L. D. Gennaro and M. Ferrara, "Sleep spindles: An overview", *Sleep Medicine Reviews*, vol. 7, no. 5, pp. 423 –440, 2003, ISSN: 1087-0792. DOI: 10.1053/smrv.2002.0252.

[46] V Gerla, "Automated analysis of long-term eeg signals", PhD thesis, Czech Technical University in Prague, Prague, Czech Republic, 2012.

[47] V. Gerla, V. Kremen, N. Covassin, L. Lhotska, E. Saifutdinova, J. Bukartyk, V. Marik, and V. Somers, "Automatic identification of artifacts and unwanted physiologic signals in eeg and eog during wakefulness", *Biomedical Signal Processing and Control*, vol. 31, pp. 381 –390, 2017, ISSN: 1746-8094. DOI: 10.1016/j.bspc.2016.09.006.

[48] V. Gerla, V. Kremen, M. Macas, E. Saifutdinova, A. Mladek, and L. Lhotska, "Expert-in-the-loop learning for sleep eeg data", in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Dec. 2018, pp. 2590–2596. DOI: 10.1109/BIBM.2018.8621557.

[49] V. Gerla, M. Murgas, A. Mladek, E. Saifutdinova, M. Macas, and L. Lhotska, "Hybrid hierarchical clustering algorithm used for large datasets: A pilot study on long-term sleep data", in *Precision Medicine Powered by pHealth and Connected Health*, N. Maglaveras, I. Chouvarda, and P. de Carvalho, Eds., Singapore: Springer Singapore, 2018, pp. 3–7, ISBN: 978-981-10-7419-6.

[50] V. Gerla, E. Saifutdinova, M. Macas, A. Mladek, and L. Lhotska, "P01-comparison of short-time fourier transform and continuous wavelet transform for frequency analysis of sleep eeg", *Clinical Neurophysiology*, vol. 129, no. 4, e14, 2018, ISSN: 1388-2457. DOI: 10.1016/j.clinph.2018.01.046.

[51] V. Gerla, E. Saifutdinova, M. Macaš, A. Mládek, and L. Lhotská, "Comparison of short-time fourier transform and continuous wavelet transform for frequency analysis of sleep eeg", in *64. SPOLEČNÝ SJEZD ČESKÉ A SLOVENSKÉ SPOLEČNOSTI PRO KLINICKOU NEUROFYZIOLOGII.*, Sep. 2017, pp. 26–27.

[52] V. Gerla, E. A. Saifutdinova, V. Kremen, M. Huptych, V. Krajca, and L. Lhotska, "A fully unsupervised clustering on adaptively segmented long-term eeg data", in *World Congress on Medical Physics and Biomedical Engineering, June 7-12, 2015, Toronto, Canada*, D. A. Jaffray, Ed., Cham: Springer International Publishing, 2015, pp. 1002–1005, ISBN: 978-3-319-19387-8.

[53] A. Gevins and A. Rémond, *Methods of analysis of brain electrical and magnetic signals*, ser. Handbook of electroencephalography and clinical neurophysiology. Elsevier, 1987, ISBN: 9780444808042. [Online]. Available: https://books.google.cz/books?id=6gsiAQAAMAAJ.

[54] E. Goldensohn, A. Legatt, S. Koszer, and S. Wolf, *Goldensohn's eeg interpretation: Problems of overreading and underreading*. Wiley, 1999, ISBN: 9780879930868. [Online]. Available: https://books.google.cz/books?id=F_qOQgAACAAJ.

[55] E. Huupponen, G. Gómez-Herrero, A. Saastamoinen, A. Värri, J. Hasan, and S.-L. Himanen, "Development and comparison of four sleep spindle detection methods", *Artificial Intelligence in Medicine*, vol. 40, no. 3, pp. 157 –170, 2007, ISSN: 0933-3657. DOI: 10.1016/j.artmed.2007.04.003.

[56] C. Iber and A. A. of Sleep Medicine, *The aasm manual for the scoring of sleep and associated events: Rules, terminology and technical specifications*. American Academy of Sleep Medicine, 2007.

[57] C. Im, *Computational eeg analysis: Methods and applications*, ser. Biological and Medical Physics, Biomedical Engineering. Springer Singapore, 2018, ISBN: 9789811309083. [Online]. Available: https://books.google.de/books?id=DztpDwAAQBAJ.

[58] A. Iranzo, B. Frauscher, H. Santos, V. Gschliesser, L. Ratti, T. Falkenstetter, C. Stürner, M. Salamero, E. Tolosa, W. Poewe, J. Santamaria, and B. Högl, "Usefulness of the sinbar electromyographic montage to detect the motor and vocal manifestations occurring in rem sleep behavior disorder", *Sleep Medicine*, vol. 12, no. 3, pp. 284 –288, 2011, ISSN: 1389-9457. DOI: https://doi.org/10.1016/j.sleep.2010.04.021. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389945710003813.

[59] M. K. Islam, A. Rastegarnia, and Z. Yang, "Methods for artifact detection and removal from scalp eeg: A review", *Neurophysiologie Clinique/Clinical Neurophysiology*, vol. 46, no. 4, pp. 287 –305, 2016, ISSN: 0987-7053. DOI: 10.1016/j.neucli.2016.07.002.

[60] M. M. Kabir, R. Tafreshi, D. B. Boivin, and N. Haddad, "Enhanced automated sleep spindle detection algorithm based on synchrosqueezing", *Medical & Biological Engineering & Computing*, vol. 53, pp. 635–644, Jul. 2015, ISSN: 1741-0444. DOI: 10.1007/s11517-015-1265-z.

[61] J. F. Kaiser, "On a simple algorithm to calculate the 'energy', of a signal", in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 1, Albuquerque, NM, 1990, pp. 381–384.

[62] A. Kantchelian, M. C. Tschantz, S. Afroz, B. Miller, V. Shankar, R. Bachwani, A. D. Joseph, and J. D. Tygar, "Better malware ground truth: Techniques for weighting anti-virus vendor labels", in *Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security, Denver, Colorado, USA*, ser. AISec '15, ACM, 2015, pp. 45–56, ISBN: 978-1-4503-3826-4. DOI: 10.1145/2808769.2808780.

[63] M. A. Klados, C. Papadelis, C. Braun, and P. D. Bamidis, "Reg-ica: A hybrid methodology combining blind source separation and regression techniques for the rejection of ocular artifacts", *Biomedical Signal Processing and Control*, vol. 6, no. 3, pp. 291 –300, 2011, ITAB 2009, ISSN: 1746-8094. DOI: 10.1016/j.bspc.2011.02.001.

[64] V. Knoblauch, M. Münch, K. Blatter, W. L. J. Martens, C. Schröder, C. Schnitzler, A. Wirz-Justice, and C. Cajochen, "Age-Related Changes in the Circadian Modulation of Sleep-Spindle Frequency During Nap Sleep", *Sleep*, vol. 28, no. 9, pp. 1093–1101, Sep. 2005, ISSN: 0161-8105. DOI: 10.1093/sleep/28.9.1093.

[65] M. Kryger, T. Roth, and W. Dement, *Principles and practice of sleep medicine. 5th edition.* Elsevier, 2005, ISBN: 9781437736090.

[66] E. Kvedalen, "Signal processing using the teager energy operator and other nonlinear operators", PhD thesis, University of Oslo, Oslo, Narway, 2003.

[67] D. Lachner-Piza, N. Epitashvili, A. Schulze-Bonhage, T. Stieglitz, J. Jacobs, and M. Dümpelmann, "A single channel sleep-spindle detector based on multivariate classification of eeg epochs: Mussdet", *Journal of Neuroscience Methods*, vol. 297, pp. 31 –43, 2018, ISSN: 0165-0270. DOI: 10.1016/j.jneumeth.2017.12.023.

[68] K. Lacourse, J. Delfrate, J. Beaudry, P. Peppard, and S. C. Warby, "A sleep spindle detection algorithm that emulates human expert spindle scoring", *Journal of Neuroscience Methods*, vol. 316, pp. 3 –11, 2019, Methods and models in sleep research: A Tribute to Vincenzo Crunelli, ISSN: 0165-0270. DOI: 10.1016/j.jneumeth.2018.08.014.

[69] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data.", *Biometrics*, vol. 33, pp. 159–174, 1977.

[70] J. Levitt, A. Nitenson, S. Koyama, L. Heijmans, J. Curry, J. T. Ross, S. Kamerling, and C. Y. Saab, "Automated detection of electroencephalography artifacts in human, rodent and canine subjects using machine learning", *Journal of Neuroscience Methods*, vol. 307, pp. 53 –59, 2018, ISSN: 0165-0270. DOI: 10.1016/j.jneumeth.2018.06.014.

[71] Y. Li, K. M. Wong, and H. D. Bruin, "Electroencephalogram signals classification for sleepstate decision - a riemannian geometry approach", *IET Signal Processing*, vol. 6, no. 4, pp. 288–299, Jun. 2012, ISSN: 1751-9675. DOI: 10.1049/iet-spr.2011.0234.

[72] E Limoges, L Mottron, C Bolduc, C Berthiaume, and R Godbout, "Atypical sleep architecture and the autism phenotype", *Brain : A journal of neurology*, vol. 128, no. 5, pp. 1049–61, May 2005.

[73] M.-Y. Liu, A. Huang, and N. E. Huang, "Evaluating and improving automatic sleep spindle detection by using multi-objective evolutionary algorithms", *Frontiers in Human Neuroscience*, vol. 11, p. 261, 2017, ISSN: 1662-5161. DOI: 10.3389/fnhum.2017.00261.

[74] F Lotte, M Congedo, A Lecuyer, F Lamarche, and B Arnaldi, "A review of classification algorithms for eeg-based brain-computer interfaces", *Journal of Neural Engineering*, vol. 4, no. 2, R1, 2007. [Online]. Available: http://stacks.iop.org/1741-2552/4/i=2/a=R01.

[75] G. Luca, J. H. Rubio, D. Andries, N. Tobback, P. Vollenweider, G. Waeber, P. M. Vidal, M. Preisig, R. Heinzer, and M. Tafti, "Age and gender variations of sleep in subjects without sleep disorders", *Annals of Medicine*, vol. 47, no. 6, pp. 482–491, 2015, PMID: 26224201. DOI: 10.3109/07853890.2015.1074271. eprint: https://doi.org/10.3109/07853890.2015.1074271. [Online]. Available: https://doi.org/10.3109/07853890.2015.1074271.

[76] M. Macas, N. Grimova, V. Gerla, L. Lhotska, and E. Saifutdinova, "Active learning for semiautomatic sleep staging and transitional eeg segments", in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Dec. 2018, pp. 2621–2627. DOI: 10.1109/BIBM.2018.8621339.

[77] M. W. Mahowald and C. H. Schenck, "Insights from studying human sleep disorders", *Nature*, vol. 437, 1279–1285, 2005, ISSN: 1476-4687. DOI: 10.1038/nature04287.

[78] A. Malafeev, X. Omlin, A. Wierzbicka, A. Wichniak, W. Jernajczyk, R. Riener, and P. Achermann, "Automatic artefact detection in single-channel sleep eeg recordings", *Journal of Sleep Research*, vol. 0, no. 0, e12679, 2018. DOI: 10.1111/jsr.12679.

[79] J. Malmivuo and R. Plonsey, *Bioelectromagnetism : Principles and applications of bioelectric and biomagnetic fields*. New York: Oxford University Press, 1995, ISBN: 0-19-505823-2.

[80] N. Martin, M. Lafortune, J. Godbout, M. Barakat, R. Robillard, G. Poirier, C. Bastien, and J. Carrier, "Topography of age-related changes in sleep spindles", *Neurobiology of Aging*, vol. 34, no. 2, pp. 468 –476, 2013, ISSN: 0197-4580. DOI: 10.1016/j.neurobiolaging.2012.05.020.

[81] L. Mayaud, S. Cabanilles, A. V. Langhenhove, M. Congedo, A. Barachant, S. Pouplin, S. Filipe, L. Pétégnief, O. Rochecouste, E. Azabou, C. Hugeron, M. Lejaille, D. Orlikowski, and D. Annane, "Brain-computer interface for the communication of acute patients: A feasibility study and a randomized controlled trial comparing performance with healthy participants and a traditional assistive device", *Brain-Computer Interfaces*, vol. 3, no. 4, pp. 197–215, 2016. DOI: 10.1080/2326263X.2016.1254403.

[82] A. Mayeli, V. Zotev, H. Refai, and J. Bodurka, "Real-time eeg artifact correction during fmri using ica", *Journal of Neuroscience Methods*, vol. 274, pp. 27 –37, 2016, ISSN: 0165-0270. DOI: 10.1016/j.jneumeth.2016.09.012.

[83] H. Merica, R. Blois, and J. M. Gaillard, "Spectral characteristics of sleep eeg in chronic insomnia", *European Journal of Neuroscience*, vol. 10, no. 5, pp. 1826–1834, 1998. DOI: 10.1046/j.1460-9568.1998.00189.x.

[84] I. Milerska, V. Kremen, V. Gerla, E. K. S. Louis, and L. Lhotska, "Semi-automated detection of polysomnographic rem sleep without atonia (rswa) in rem sleep behavioral disorder", *Biomedical Signal Processing and Control*, vol. 51, pp. 243 –252, 2019, ISSN: 1746-8094. DOI: https://doi.org/10.1016/j.bspc.2019.02.023. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1746809419300679.

[85] H Moldofsky, F. A. Lue, J. R. Davidson, and R Gorczynski, "Effects of sleep deprivation on human immune functions.", *The FASEB Journal*, vol. 3, no. 8, pp. 1972–1977, 1989, PMID: 2785942. DOI: 10.1096/fasebj.3.8.2785942.

[86] M. K. I. Molla, M. R. Islam, T. Tanaka, and T. M. Rutkowski, "Artifact suppression from eeg signals using data adaptive time domain filtering", *Neurocomputing*, vol. 97, pp. 297 –308, 2012, ISSN: 0925-2312. DOI: 10.1016/j.neucom.2012.05.009.

[87] M. Molle, "Grouping of spindle activity during slow oscillations in human non-rapid eye movement sleep", *The Journal of Neuroscience*, pp. 10 941–47, 2002.

[88] M. Mölle, L. Marshall, S. Gais, and J. Born, "Grouping of spindle activity during slow oscillations in human non-rapid eye movement sleep", *Journal of Neuroscience*, vol. 22, no. 24, pp. 10 941–10 947, 2002, ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.22-24-10941.2002.

[89] I Myatchin and L Lagae, "Sleep spindle abnormalities in children with generalized spike-wave discharges", *Pediatric Neurology*, vol. 36, no. 2, pp. 106–11, Mar. 2007.

[90] X. Navarro-Sune, A. L. Hudson, F. D. V. Fallani, J. Martinerie, A. Witon, P. Pouget, M. Raux, T. Similowski, and M. Chavez, "Riemannian geometry applied to detection of respiratory states from eeg signals: The basis for a brain–ventilator interface", *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 5, pp. 1138–1148, May 2017, ISSN: 0018-9294. DOI: 10.1109/TBME.2016.2592820.

[91] A. Ng, *Cs229 lecture notes*. Simon Fraser University, 2004.

[92] H.-A. T. Nguyen, J. Musson, F. Li, W. Wang, G. Zhang, R. Xu, C. Richey, T. Schnell, F. D. McKenzie, and J. Li, "Eog artifact removal using a wavelet neural network", *Neurocomputing*, vol. 97, pp. 374 –389, 2012, ISSN: 0925-2312. DOI: 10.1016/j.neucom.2012.04.016.

[93] H. Nolan, R. Whelan, and R. Reilly, "Faster: Fully automated statistical thresholding for eeg artifact rejection", *Journal of Neuroscience Methods*, vol. 192, no. 1, pp. 152 –162, 2010, ISSN: 0165-0270. DOI: http://dx.doi.org/10.1016/j.jneumeth.2010.07.015.

[94] C. O'Reilly, N. Gosselin, J. Carrier, and T. Nielsen, "Montreal archive of sleep studies: An open-access resource for instrument benchmarking and exploratory research", *Journal of Sleep Research*, vol. 23, no. 6, pp. 628–635, 2014. DOI: 10.1111/jsr.12169. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jsr.12169. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/jsr.12169.

[95] C. O'Reilly and T. Nielsen, "Automatic sleep spindle detection: Benchmarking with fine temporal resolution using open science tools", *Frontiers in Human Neuroscience*, vol. 9, p. 353, 2015, ISSN: 1662-5161. DOI: 10.3389/fnhum.2015.00353.

[96] A. Parekh, I. W. Selesnick, R. S. Osorio, A. W. Varga, D. M. Rapoport, and I. Ayappa, "Multichannel sleep spindle detection using sparse low-rank optimization", *Journal of Neuroscience Methods*, vol. 288, pp. 1 –16, 2017, ISSN: 0165-0270. DOI: 10.1016/j.jneumeth.2017.06.004.

[97] C. R. Patti, T. Penzel, and D. Cvetkovic, "Sleep spindle detection using multivariate gaussian mixture models", *Journal of Sleep Research*, vol. 27, no. 4, e12614, 2018. DOI: 10.1111/jsr.12614.

[98] J. Perrier, P. Clochon, F. Bertran, C. Couque, J. Bulla, P. Denise, and M.-L. Bocca, "Specific eeg sleep pattern in the prefrontal cortex in primary insomnia", *PLOS ONE*, vol. 10, no. 1, pp. 1–14, Jan. 2015. DOI: 10.1371/journal.pone.0116864.

[99] D. Petit, J.-F. Gagnon, M. L. Fantini, L. Ferini-Strambi, and J. Montplaisir, "Sleep and quantitative eeg in neurodegenerative disorders", *Journal of Psychosomatic Research*, vol. 56, no. 5, pp. 487 –496, 2004, ISSN: 0022-3999. DOI: 10.1016/j.jpsychores.2004.02.001.

[100] P.-L. Ratti, N. Amato, O. David, and M. Manconi, "A high-density polysomnographic picture of disorders of arousal", *Sleep*, vol. 41, no. 11, zsy162, 2018. DOI: 10.1093/sleep/zsy162.

[101] V. C. Raykar, S. Yu, L. H. Zhao, A. Jerebko, C. Florin, G. H. Valadez, L. Bogoni, and L. Moy, "Supervised learning from multiple experts: Whom to trust when everyone lies a bit", in *Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, Quebec, Canada*, ser. ICML '09, ACM, 2009, pp. 889–896, ISBN: 978-1-60558-516-1. DOI: 10.1145/1553374.1553488.

[102] A. Rechtschaffen and A. Kales, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects", 1968.

[103] S. Redline, H. L. Kirchner, S. F. Quan, D. J. Gottlieb, V. Kapur, and A. Newman, "The Effects of Age, Sex, Ethnicity, and Sleep-Disordered Breathing on Sleep Architecture", *JAMA Internal Medicine*, vol. 164, no. 4, pp. 406–418, Feb. 2004, ISSN: 2168-6106. DOI: 10.1001/archinte.164.4.406. eprint: https://jamanetwork.com/journals/jamainternalmedicine/articlepdf/216692/ioi20991.pdf. [Online]. Available: https://doi.org/10.1001/archinte.164.4.406.

[104] S. Reutrakul, T. Anothaisintawee, S. J. Herring, B. I. Balserak, I. Marc, and A. Thakkinstian, "Short sleep duration and hyperglycemia in pregnancy: Aggregate and individual patient data meta-analysis", *Sleep Medicine Reviews*, vol. 40, pp. 31 –42, 2018, ISSN: 1087-0792. DOI: https://doi.org/10.1016/j.smrv.2017.09.003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S108707921630140X.

[105] C. Reynolds, M. Short, and M. Gradisar, "Sleep spindles and cognitive performance across adolescence: A meta-analytic review", *Journal of Adolescence*, vol. 66, pp. 55 –70, 2018, ISSN: 0140-1971. DOI: 10.1016/j.adolescence.2018.04.003.

[106] P. L. C. Rodrigues, F. Bouchard, M. Congedo, and C. Jutten, "Dimensionality Reduction for BCI classification using Riemannian geometry", in *7TH GRAZ BRAIN-COMPUTER INTERFACE CONFERENCE (BCI 2017)*, Gernot R. Müller-Putz, Graz, Austria, Sep. 2017. DOI: 10.3217/978-3-85125-533-1-16.

[107] E. Saifutdinova, D. U. Dudysová, V. Gerla, and L. Lhotská, "Improvement sleep spindle detection by aggregation techniques", in *15th Mediterranean Conference on Medical and Biological Engineering and Computing (MEDICON2019)*, 2019.

[108] E. Saifutdinova, D. U. Dudysová, L. Lhotská, V. Gerla, and M. Macaš, "Artifact detection in multichannel sleep eeg using random forest classifier", in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Dec. 2018, pp. 2803–2805. DOI: 10.1109/BIBM.2018.8621374.

[109] E. Saifutdinova, V. Gerla, and L. Lhotska, "Adaptive segmentation with successive windows", in *Proceedings of BioDat 2014 - Conference on Advanced Methods of Biological Data and Signal Processing*, 2014.

[111] E. Saifutdinova, V. Gerla, M. Macas, and L. Lhotska, "P04-spatial geometric analysis in sleep polysomnographic data", *Clinical Neurophysiology*, vol. 129, no. 4, e15, 2018, ISSN: 1388-2457. DOI: 10.1016/j.clinph.2018.01.049.

[112] E. Saifutdinova, V. Gerla, M. Macaš, and L. Lhotská, "Spatial geometric analysis in sleep polysomnographic data", in *64. SPOLEČNÝ SJEZD ČESKÉ A SLOVENSKÉ SPOLEČNOSTI PRO KLINICKOU NEUROFYZIOLOGII.*, Sep. 2017, pp. 26–27.

[113] E. Saifutdinova, M. Congedo, D. Dudysova, L. Lhotska, J. Koprivova, and V. Gerla, "An unsupervised multichannel artifact detection method for sleep eeg based on riemannian geometry", *Sensors*, vol. 19, no. 3, 2019, ISSN: 1424-8220. DOI: 10.3390/s19030602.

[114] E. Saifutdinova, V. Gerla, and L. Lhotská, "Riemannian geometry in sleep stage classification", in *Information Technology in Bio- and Medical Informatics*, M. Bursa, A. Holzinger, M. E. Renda, and S. Khuri, Eds., Cham: Springer International Publishing, 2017, pp. 92–99, ISBN: 978-3-319-64265-9.

[115] E. Saifutdinova, J. Koprivova, L. Lhotska, and M. Macas, "Topological properties of functional brain connectivity in obsessive-compulsive disorder", in *XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016*, E. Kyriacou, S. Christofides, and C. S. Pattichis, Eds., Cham: Springer International Publishing, 2016, pp. 157–161, ISBN: 978-3-319-32703-7.

[117] C. B. Saper, T. E. Scammell, and J. Lu, "Hypothalamic regulation of sleep and circadian rhythms", *Nature*, vol. 437, no. 8, pp. 1257–1263, 7063 2005. DOI: 10.1038/nature04284.

[118] M. J. Sateia, "International classification of sleep disorders-third edition", *Chest*, vol. 146, no. 5, pp. 1387 –1394, 2014, ISSN: 0012-3692. DOI: 10.1378/chest.14-0970.

[119] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "Dbscan revisited, revisited: Why and how you should (still) use dbscan", *ACM Trans. Database Syst.*, vol. 42, no. 3, 19:1–19:21, Jul. 2017, ISSN: 0362-5915. DOI: 10.1145/3068335. [Online]. Available: http://doi.acm.org/10.1145/3068335.

[120] F. Siclari, G. Bernardi, J. Cataldi, and G. Tononi, "Dreaming in nrem sleep: A high-density eeg study of slow waves and spindles", *Journal of Neuroscience*, vol. 38, no. 43, pp. 9175–9185, 2018, ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.0855-18.2018.

[121] J. Stern and J. Engel, *Atlas of eeg patterns*, ser. LWW medical book collection. Lippincott Williams & Wilkins, 2005, ISBN: 9780781741248. [Online]. Available: https://books.google.cz/books?id=n6HxBvzEYaMC.

[122] A. Szentkirályi, K. K. Wong, R. R. Grunstein, A. L. D'Rozario, and J. W. Kim, "Performance of an automated algorithm to process artefacts for quantitative eeg analysis during a simultaneous driving simulator performance task", *International Journal of Psychophysiology*, vol. 121, pp. 12 –17, 2017, ISSN: 0167-8760. DOI: 10.1016/j.ijpsycho.2017.08.004.

[123] M. Taherisadr, O. Dehzangi, and H. Parsaei, "Single channel eeg artifact identification using two-dimensional multi-resolution analysis", *Sensors*, vol. 17, no. 12, 2017, ISSN: 1424-8220. DOI: 10.3390/s17122895.

[124] F. Taya, Y. Sun, G. Borghini, P. Aricò, F. Babiloni, A. Bezerianos, and N. V. Thakor, "Training-induced changes in information transfer efficiency of the brain network: A functional connectome approach", in *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, Apr. 2015, pp. 1028–1031. DOI: 10.1109/NER.2015.7146802.

[125] A Varri, "Digital processing of the eeg in epilepsy", PhD thesis, Tampere University of Technology, Tampere, Finland, 1988.

[126] K. Veldová, D. U. Dudysová, E. Saifutdinová, and J. Kopřivová, "Comparison of sleep characteristics in chronic insomnia subtypes: Eeg correlates of sleep misperception", *Sleep Medicine*, vol. 40, e337 –e338, 2017, ISSN: 1389-9457. DOI: 10.1016/j.sleep.2017.11.992.

[127] E. M. Ventouras, E. A. Monoyiou, P. Y. Ktonas, T. Paparrigopoulos, D. G. Dikeos, N. K. Uzunoglu, and C. R. Soldatos, "Sleep spindle detection using artificial neural networks trained with filtered time-domain eeg: A feasibility study", *Computer Methods and Programs in Biomedicine*, vol. 78, no. 3, pp. 191 –207, 2005, ISSN: 0169-2607. DOI: 10.1016/j.cmpb.2005.02.006.

[128] D. C. Wallant, V. Muto, G. Gaggioni, M. Jaspar, S. L. Chellappa, C. Meyer, G. Vandewalle, P. Maquet, and C. Phillips, "Automatic artifacts and arousals detection in whole-night sleep eeg recordings", *Journal of Neuroscience Methods*, vol. 258, pp. 124 –133, 2016, ISSN: 0165-0270. DOI: 10.1016/j.jneumeth.2015.11.005.

[129] E. J. Wamsley and et al., "Reduced sleep spindles and spindle coherence in schizophrenia: Mechanisms of impaired memory consolidation?", *Biological Psychiatry*, vol. 71, no. 2, pp. 154–61, Jan. 2012.

[130] E. J. Wamsley, M. A. Tucker, A. K. Shinn, K. E. Ono, S. K. McKinley, A. V. Ely, D. C. Goff, R. Stickgold, and D. S. Manoach, "Reduced sleep spindles and spindle coherence in schizophrenia: Mechanisms of impaired memory consolidation?", *Biological Psychiatry*, vol. 71, no. 2, pp. 154 –161, 2012, Functional Consequences of Altered Cortical Development in Schizophrenia, ISSN: 0006-3223. DOI: 10.1016/j.biopsych.2011.08.008.

[131] S. C. Warby, S. L. Wendt, P. Welinder, E. G. Munk, O. Carrillo, H. B. D. Sorensen, P. Jennum, P. E. Peppard, P. Perona, and E. Mignot, "Sleep spindle detection: Crowdsourcing and evaluating performance of experts, non-experts, and automated methods", in *Nature Methods*, 2014.

[132] C. Yücelbaş, Ş. Yücelbaş, S. Özşen, G. Tezel, S. Küççüktürk, and Ş. Yosunkaya, "Automatic detection of sleep spindles with the use of stft, emd and dwt methods", *Neural Computing and Applications*, vol. 29, no. 8, pp. 17–33, Apr. 2018, ISSN: 1433-3058. DOI: 10.1007/s00521-016-2445-y.

[133] M. Šmotek, P. Vlček, E. Saifutdinova, and J. Kopřivová, "Does blue-light increase cortical activation? findings from an eeg study", *Sleep Medicine*, vol. 40, e310 – e311, 2017, ISSN: 1389-9457. DOI: 10.1016/j.sleep.2017.11.914.