Bachelor Project

Czech Technical University in Prague

F3

Faculty of Electrical Engineering
Department of Cybernetics

# Artefact detection in multichannel sleep PSG

Kudryakova Maria

Supervisor: Mgr. Elizaveta Saifutdinova
May 2018

# BACHELOR'S THESIS ASSIGNMENT

## I. Personal and study details

Student's name: **Kudryakova Maria** Personal ID number: **453514**

Faculty / Institute: **Faculty of Electrical Engineering**

Department / Institute: **Department of Cybernetics**

Study program: **Cybernetics and Robotics**

Branch of study: **Robotics**

## II. Bachelor's thesis details

Bachelor's thesis title in English:

**Artefact Detection in Multichannel Sleep PSG**

Bachelor's thesis title in Czech:

**Detekce artefaktů v multikanálovém spánkovém PSG**

Guidelines:

1. Study the problem of automatic artefact detection in sleep PSG. Describe methods used for detection of different types of artefacts.
2. Analyze artefact types presented in the open-source DREAMS artifacts database. Select suitable features what could be used in automatic artefact detection. Implement state-of-the art method for artefact detection.
3. Implement artifact detection method in multichannel sleep PSG using adaptive segmentation based on Riemannian geometry and suggested machine learning techniques.
4.Compare and evaluate obtained results for the implemented and state-of-the-art methods. Present and evaluate results for different types of artefacts and used channels

Bibliography / sources:

[1] Sanei. S, Jonathon C. - EEG signal processing - 2009. Chichester, England: John Wiley & Sons
[2] Motamedi-Fakhr, S., Moshrefi-Torbati, M., Hill, M., Hill, C. and White, P - Signal processing techniques applied to human sleep EEG signals -A review. Biomedical Signal Processing and Control. - 2014
[3] Nolan, H., Whelan, R. and Reilly, R. - FASTER: Fully Automated Statistical Thresholding for EEG artifact Rejection. Journal of Neuroscience Methods - 2010
[4] Ing. Václav Gerla - Automated Analysis of Long-Term EEG Signals. Doctoral Thesis - 2012. Prague

Name and workplace of bachelor's thesis supervisor:

**Mgr. Elizaveta Saifutdinova, Department of Cybernetics, FEE**

Name and workplace of second bachelor's thesis supervisor or consultant:

Date of bachelor's thesis assignment: **09.01.2018** Deadline for bachelor thesis submission: **25.05.2018**

Assignment valid until: **30.09.2019**

| Mgr. Elizaveta Saifutdinova | doc. Ing. Tomáš Svoboda, Ph.D. | prof. Ing. Pavel Ripka, CSc. |
|---|---|---|
| Supervisor's signature | Head of department's signature | Dean's signature |

## III. Assignment receipt

The student acknowledges that the bachelor's thesis is an individual work. The student must produce her thesis without the assistance of others, with the exception of provided consultations. Within the bachelor's thesis, the author must state the names of consultants and include a list of references.

. | Date of assignment receipt | Student's signature

# Acknowledgements

I want to thank my supervisor Elizaveta Saifutdinova for her guidance, constant supervision, providing all necessary information needed for the project and also support in completing the project.


# Declaration

I declare that the presented work was developed independently and that I have listed all sources of information used within it in accordance with the methodical instructions for observing the ethical principles in the preparation of university theses.
Prague, date

Prohlašuji, že jsem předloženou práci vypracovala samostatně, a že jsem uvedla veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.
V Praze, dne

# Abstract

A method of automatic artefact detection in sleep PSG is proposed in this work. It is based on classification of segments with different lengths. A new multichannel adaptive segmentation was proposed. For comparison of multichannel data, Riemannian distance was used. Classification was performed with Naïve Bayes classifier. The method was tested on open-source the DREAMS Artefacts Database. The detection method was evaluated by various statistical metrics. They were compared with results provided by Stephanie Devuyst, the author of the database, and detection based on classification of segments with constant length. The results show increase in all metrics. Mainly, $F_1$ score is higher on 30 % in comparison with method performed by Devuyst and on 20 % comparing to the state of art method.



**Keywords:** PSG, artefacts, adaptive segmentation, Riemannian distance, Naïve Bayes classifier



**Supervisor:** Mgr. Elizaveta Saifutdinova
Department of Cybernetics
Faculty of Electrical Engineering
Czech Technical University in Prague
Technická 2
160 00 Prague 6
Czech Republic



# Abstrakt

V této práci je navržena metoda automatické detekce artefaktu ve spánkovém PSG, která je založená na klasifikaci segmentů s různými délkami.V rámci ní navržen nový způsob multikanálové adaptivní segmentace. Pro porovnání multikanálových dat byla použita Riemannova vzdálenost. Klasifikace byla provedena s použitím naivního bayesovského klasifikátoru. Metoda byla testována na open-source datech z DREAMS Artefacts Database. Klasifikace byla hodnocena pomocí různých statistických metrik. Výsledky metody byly porovnávány s výsledky, které poskytuje autor databáze Stephanie Devuyst, a detekce na základě klasifikace segmentů s konstantní délkou. Výsledky ukazují významné zvýšení ve všech metrikách. Zejména $F_1$ je vyšší o 30 % ve srovnání s metodou prováděnou Devuyst a o 20 % ve srovnání s nejmodernější metodou.



**Klíčová slova:** PSG, artefakty, adaptivní segmentace, Riemannova vzdálenost, Naivní Bayesovský klasifikátor

**Překlad názvu:** Detekce artefaktů v multikanálovém spánkovém PSG

# Contents

# Figures

# Tables

# Chapter 1

# Introduction

## 1.1 Motivation and aim of the thesis

Sleep is a vital part of our life. Every person spends a third of his or her life by sleeping. Our body and brain need it as much as oxygen, water and food. Sleep quality affects our cognitive and learning capabilities, physical health. It is one of the most crucial items in diagnostics of mental illnesses. Polysomnography (PSG) remains to be a gold standard method of sleep analysis. It provides brain activity (EEG), heart rate (ECG) the oxygen level in blood, breathing and eye or muscle movements (EOG and EMG). Various sleep disorders are diagnosed by polysomnography test. For instance, obstructive sleep apnea, which is characterized by repetitive episodes of paused breathing during sleep, or different types of hypersomnia.

The importance of artefact detection in PSG is widely acknowledged. High artefact density makes difficulties in analysis of the obtained data and, especially, during automatic processing. Manual screening and allocation of damaged parts is almost impossible in long-term recordings and takes a lot of time. The non-stationary behaviour of artefacts makes automatic detection a difficult task [4, 5].

At the moment, many methods of automatic artefact detection have been proposed. One of them using extended Kalman filter and neural network instead of autoregressive (AR) model was examined with achieving 65 % of sensitivity for 90 % of specificity [6]. Devuyst et al. presented series of algorithms for dealing with different types of artefacts with sensitivity 83.67 % and specificity 96.47 % in 2008 [7]. In 2000 was proposed a method for removing a wide variety of artefacts from EEG records based on blind source separation by independent component analysis (ICA)[8]. However, this approach requires a large number of channel for good source separation. FASTER – a fully automated statistical thresholding method for EEG artefact rejection, which also incorporates ICA showed >90 % sensitivity and specificity for detection of contaminated channels [9]. For artefact detection can be applied to machine learning, for instance, support vector machine (SVM) classifier [10].

The aim of the study is to implement automatic artefact detection method in sleep PSG using multichannel adaptive segmentation. The proposed method

is applied to the open-source DREAMS artefacts database. Results of the detection were compared with a widely used method based on classification of data segments of constant length. In the study, two big ideas were connected: multichannel approach on pre-processing stage and machine learning for detection of contaminated data segments. New proposed adaptive segmentation extracts stationary segments by analyzing Riemannian distance between neighbour data windows. Such an approach allows working with all channels at once whereas many segmentation methods are applied to channels separately. Proposed multichannel segmentation can be applied to PSG even with the extremely big number of channels. Obtained segments are classified by Naïve Bayes. Classification results of the data divided into non-overlapped segments of different length and into segments of the same length were evaluated and compared.

## 1.2 Thesis organization

The thesis is organized as follows. In Chapter 2 is presented a brief theory of PSG and artefacts. In Chapter 3 signal processing method and database are described. Chapter 4 contains the achieved results with their analysis.

## 1.3 List of Abbreviations

- PSG Polysomnography
- EEG Electroencephalography
- ECG Electrocardiography
- EMG Electromyography
- EOG Electrooculography
- W Wakefulness stage
- N1 Non-rapid eye movement sleep stage 1
- N2 Non-rapid eye movement sleep stage 2
- N3 Non-rapid eye movement sleep stage 3
- REM Rapid eye movement sleep
- NREM Non-rapid eye movement sleep
- SPD Symmetric and positive definite matrices

# Chapter 2

# Theoretical framework

## 2.1 Polysomnography

Polysomnography is a medical test to make a continuous record during sleep of multiple physiological variables. Usually consists of:

- The EEG is obtained using electrodes placed on the surface of the head by a standardized system 10-20 (Figure 2.1). Each electrode placed on the scalp is identified with a letter. For example, C is central, F is frontal, etc. Numbers define a hemisphere or the brain: even numbers - right hemisphere, odd numbers - left, "Z" (zero) refers to the midline sagittal plane. The system's name refers to constant distances between neighbouring electrodes which are either 10 % or 20 % of the total front–back or right–left distance of the skull [11]. These electrodes are paired to create channels and the potential difference between a pair of electrodes is measured. Typically, there are presented 19 channels covering the whole head. An example is showed in the Figure 2.1.

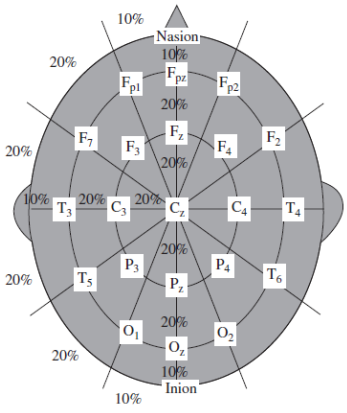



**Figure 2.1:** The EEG electrode placement according to the 10-20 system [1].

- The ECG consists of 12 electrodes (12-Lead ECG electrode placement system). It is applied for detection of cardiac events and abnormalities like tachycardia or bradycardia.

- The EOG includes 2 electrodes for eye movement detection. There are two options for the location of these electrodes. The first variant shows vertical and lateral eye movements as waveforms of opposite polarity or phase. The electrodes in this case placed outside each other canthus with the left outer canthus 1cm below the horizontal midline, and the right outer canthus 1cm above the horizontal midline and referenced to a mastoid electrode. The second variant shows lateral movements as out-of-phase waveforms and vertical movements as in-phase waveforms. For this measurement, each outer canthus electrode is placed 1cm below the horizontal midline and referenced to the midline frontal polar electrode [12].

- The EMG is used for measuring muscle activity. Chin EMG electrodes placement by the AASM: 1cm above the inferior edge of the mandible is placed midline electrode and two electrodes located 2 cm below the inferior edge of the mandible (one 2 cm right and another 2 cm left of the midline). For measuring limb movement electrodes are placed on the anterior muscles of the lower legs [13].

- Additional channels. For instance, oxyhaemoglobin saturation or oral-nasal airflow.

There are brain rhythms and waveforms which are considered as normal (Figure 2.2), they can be defined by frequency range, amplitude and place of an occurrence [1].

- Gamma rhythms (also called fast beta rhythm) has frequency above 20 Hz. The amplitudes of these rhythms are very low and their occurrence is rare.

- Alpha rhythms are characterized by frequency in the range 8-12 Hz and amplitude 5-15 $\mu$V. This waves usually appear when our eyes are closed and during relaxation.

- Beta rhythms are the usual rhythm associated with active thinking, focusing. The frequency is above 13 Hz.

- Theta rhythms have amplitude between 10-50$\mu$V and frequency 4-7.5 Hz. They usually appear during N1 and N2, the persistent occurrence of these waves in the waking adult is abnormal.

- Delta rhythms lie within the range of 0.5–4 Hz. These waves usually define deep sleep (N3) and may be present in the waking state [14].

Human sleep is divided into two classes: Rapid eye movement (REM) sleep and non-REM (NREM). Further NREM is split into 3 stages (N1, N2 and N3) according to the American Academy of Sleep Medicine (AASM) [13]. NREM and REM occur in alternating cycles, each lasting approximately 90-100 minutes, with a total of 4-6 cycles. In general, in the healthy young adult NREM sleep accounts for 75-90 % of sleep time (3-5 % N1, 50-60 %

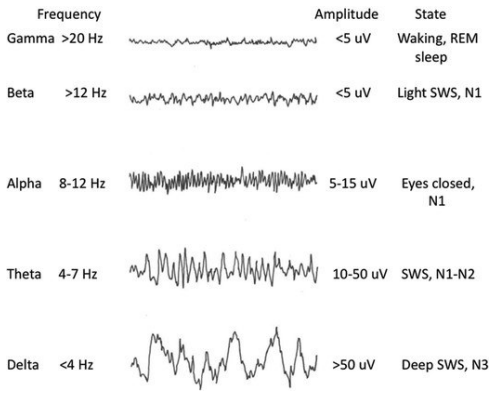



**Figure 2.2:** Samples of normal brain activity [2].

N2, and 10-20 % N3) [15]. REM sleep accounts for 10-25 % of sleep time. Sleep stages are classified by processing of the EEG, EOG and EMG signals. In the following table (Table 2.1) are presented features characterizing every stage determined by the AASM. Figure 2.3 demonstrates examples of brain activity during different sleep stages including their typical waveforms.

| Sleep Stage | EEG | EOG | EMG |
|---|---|---|---|
| W | >50% alpha rhythm | Reading or REM | Activity is normal or high |
| S1 | >50% of alpha rhythms replaced by Low Voltage Mixed Frequency (LVMF) waves, vertex sharp waves | SREM | Activity is lower than W |
| S2 | Theta waves with K-complexes and/or sleep spindles | No eye movements | Activity is lower than W |
| S3 | Slow-wave activity, sleep spindles may be present; amplitude must be at least 75 $\mu$V from peak-to-peak | No eye movements | Activity is lower than in S2 and sometimes as low as in REM. |
| REM | Theta waves, alpha waves present slower than at W; sawtooth waves (sharply contoured, triangular; often serrated (2-6 Hz) preceding bursts of rapid eye movements | REM | Significantly reduced compared to NREM sleep |

**Table 2.1:** Sleep stage characteristics.

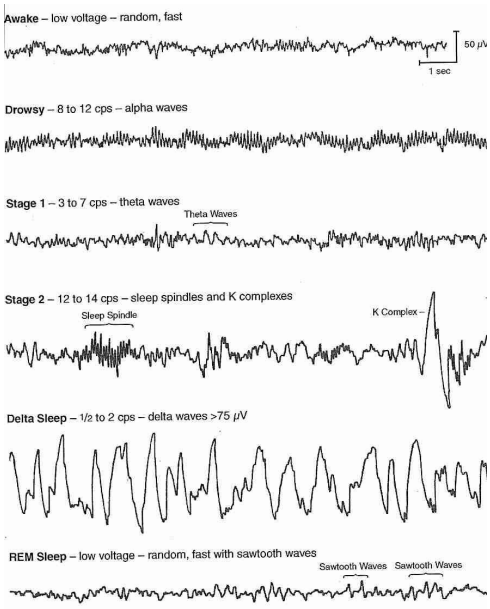



**Figure 2.3:** Samples of sleep stages and normal waveforms [3].

## 2.2 Artefacts

The PSG signals contain different artefacts. In signal processing, the artefact is a structure not normally presented, but produced by some external action; something artificial [16]. All of them can be divided into two groups: physiologic and extraphysiologic artefacts. Physiologic artefacts arise from the patient body from sources other than the brain (heartbeat, movements), when extraphysiologic comes from the external ones. After detection contaminated parts of data are removed or suppressed [4]. In this work, detection of artefacts is in the focus of interest.

EEG artefacts are characterized by atypical waveforms amplitude or frequency. For instance, slow ondulations artefact (Figure 2.4 c) has lower frequency then delta rhythm appears during sweating and deep breathing. The unusual increase of the EEG amplitude (Figure 2.4 a) is defined as instant increase and recession of amplitude, it may be caused by EOG interference. The other artefact is muscle or movement artefact (Figure 2.4 b) can be detected by a long-term increase in the amplitude of the signal. Electrode popping artefact also called abrupt transitions (Figure 2.4 d) is characterized by the occurrence of spikes. Falling electrode artefact (Figure 2.4 e,f) has amplitude close to zero[7].

The artefacts which occur in EEG channels, but can not be attributed to the mentioned types are in the Figure 2.5 a-d.

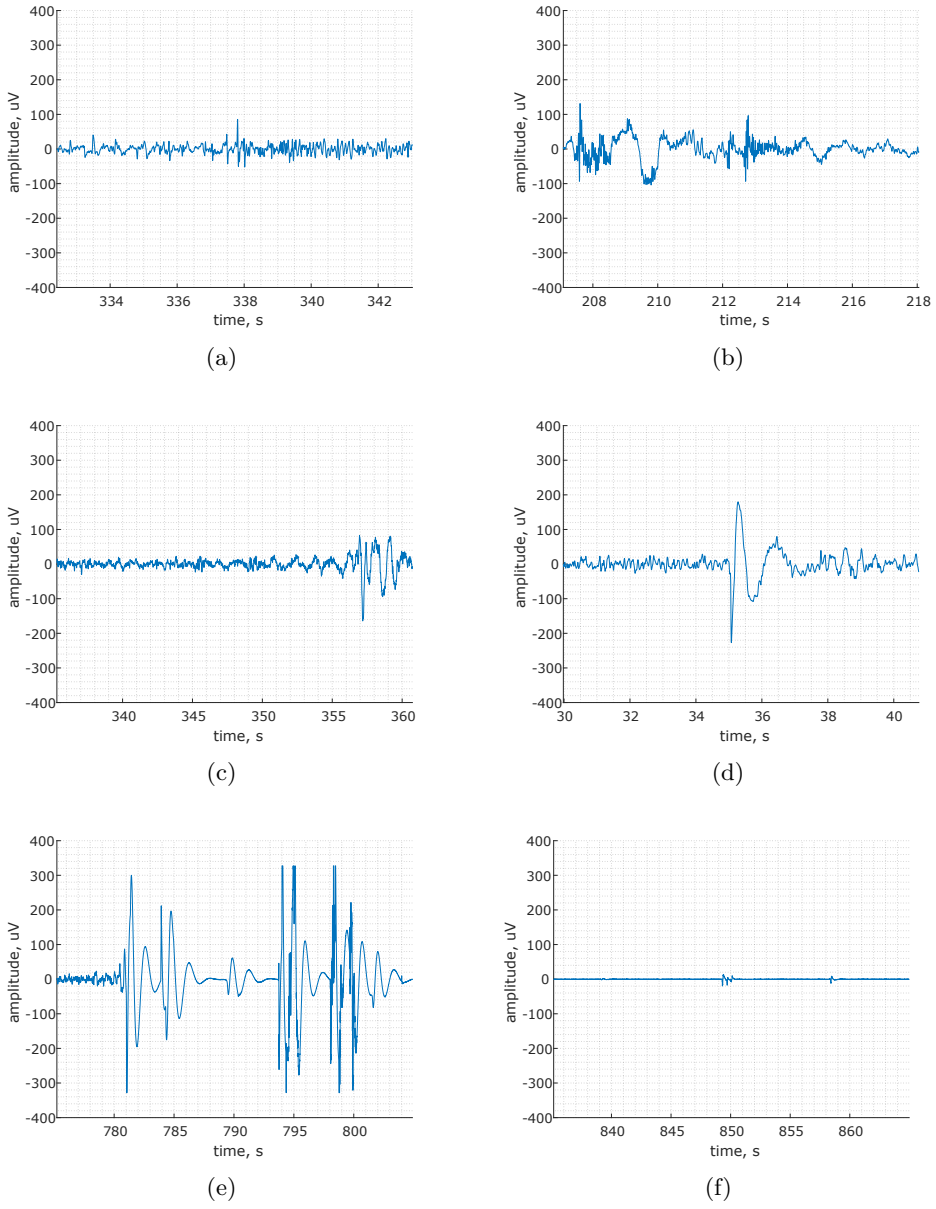



**Figure 2.4:** Typical EEG artefacts on CZ-A1 channel: unusual increase of the EEG amplitude (a), muscle or movement artefact (b), slow ondulations artefact (c), electrode popping artefact (d), falling electrode artefact (e,f).

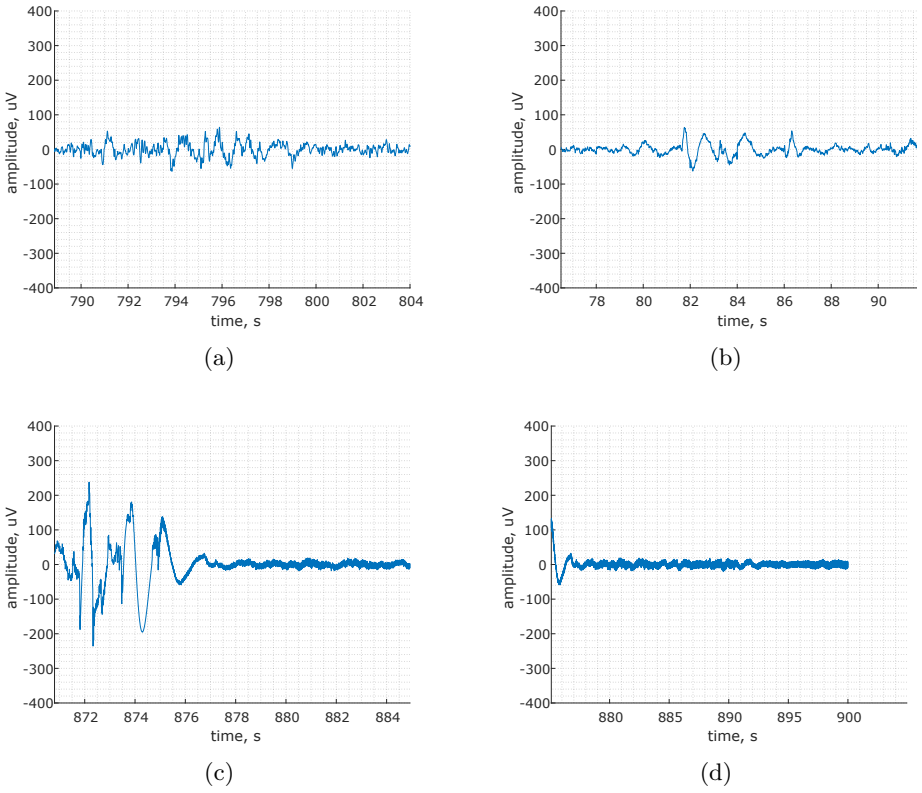



**Figure 2.5:** Other artefacts (a-d) on CZ-A1 channel.

# Chapter 3

# Material and methods

## 3.1 Database

In this work was used The DREAMS Artefacts Database which consists of data collected in a sleep laboratory using the 32-channel polygraph in a sleep laboratory in Belgium [7]. From 20 polysomnographic recording from patients with different disorders were selected those whose sampling frequency is 200 Hz. There are data of 13 subjects in total. Every recording is 15 minutes long and contains two EOG channels (P8-A1, P18-A1), three EEG channels (CZ-A1 or C3-A1, FP1-A1 and O1-A1) and submental EMG channel. Artefacts in these recordings were visually detected by clinical experts. There were detected artefacts of the following types: muscle or movement artefacts (mvtE), an unusual increase of amplitude (highE), electrode popping (transE) and falling (noE), slow ondulations (ondE) and other artefacts (otherE). Details are provided in Table 3.1.

| Subject | highE | mvtE | transE | otherE | ondE | noE | Total |
|---|---|---|---|---|---|---|---|
| 5 | 9 | 8 | 7 | - | 1 | - | 25 |
| 8 | 42 | 5 | 13 | 1 | - | 6 | 67 |
| 10 | 14 | 1 | - | - | - | - | 15 |
| 11 | - | 9 | 1 | - | - | - | 10 |
| 12 | 18 | 17 | 19 | 20 | - | - | 74 |
| 13 | 6 | 9 | - | - | 1 | - | 16 |
| 14 | 13 | 11 | 2 | - | 1 | - | 27 |
| 15 | 5 | 5 | - | 2 | - | - | 12 |
| 16 | 45 | 41 | - | 22 | 11 | - | 119 |
| 17 | 9 | 8 | - | - | - | - | 17 |
| 18 | 3 | 1 | 2 | - | - | - | 6 |
| 19 | 12 | 14 | - | - | 2 | - | 28 |
| 20 | 17 | 3 | 2 | 1 | - | - | 23 |

**Table 3.1:** Artefact detailes of recordings selected for the study. First column represents number of recording in the dataset. Last one stands for total number of artefact for every recording. In the rest columns, there are presented numbers of artefact of certain type.

## 3.2 Method overview

The applied method involves steps presented in Figure 3.1. Firstly, the data is divided into non-overlapped segments. Then, for every segment are extracted statistical and other features described in the section 3.4. Segmentation methods used in the work are presented in the section 3.3. The next step is the classification. Naive Bayes outlined in section 3.5 was chosen as a classifier. Testing of the presented method was performed by leave-one-out cross-validation. Various statistical metrics were calculated on each iteration. Details of classifier evaluation are in section 3.6 .

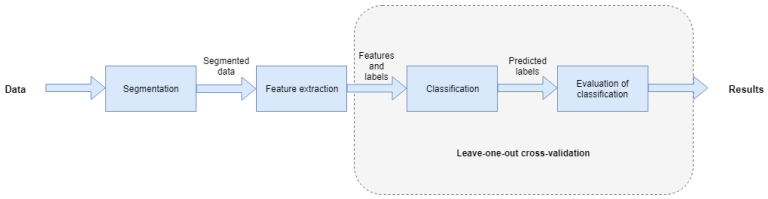



**Figure 3.1:** Method overview.

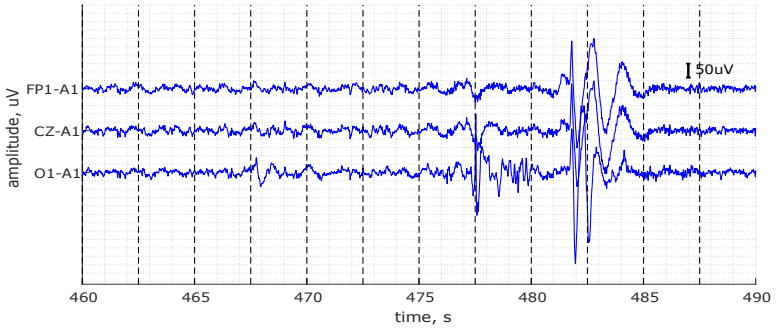
## 3.3 Segmentation

Signals are in most cases non-stationary and segmentation allows to divide them into sub-sections in which the signal is quasi-stationary. It is also important in pattern recognition, for instance, for identification of the occurrence of events in PSG recordings [4]. There are two types of segmentation: constant and adaptive.

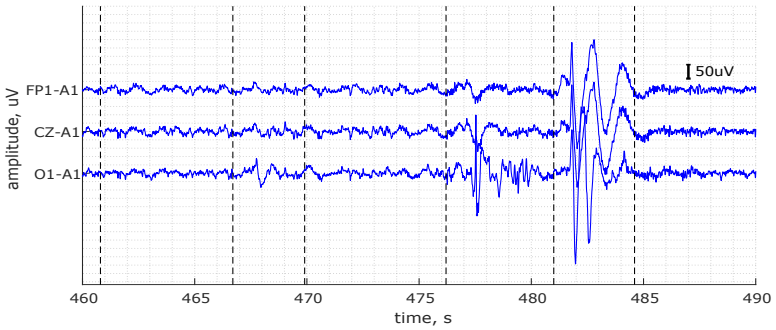
When the first approach divides the data into segments of constant length, adaptive segmentation adjusts to the signal variation and obtains segments of different length but stationary. Results of constant and adaptive segmentation are presented in Figure 3.2. As it might be observed adaptive approach returns less number of segments but each segment is highly stable in statistical properties. On the other hand, some segments obtained with constant segmentation contain features of both artefact and non-artefact classes.

### 3.3.1 Constant segmentation

Constant segmentation is the most commonly used method. A signal is divided into windows with constant duration. In multichannel case, every channel signal is divided the same. The length of the segment can vary depending on the purpose of the segmentation and the algorithm applied. Long windows can affect lose of sensitivity [17]. However, this type of segmentation is simple and stable and often used in the studies [18, 19].




(a)




(b)

**Figure 3.2:** Results of constant (a) and adaptive (b) segmentation. Segment borders are represented by dashed lines. Window length of size 2.5 s was used for constant segmentation. Parameters of adaptive segmentation are window length 2.5 s and step 10 samples.

### 3.3.2 Adaptive segmentation

Adaptive segmentation unlike to constant one divides data into segments with different length. This method is more complex but more reliable. A signal is analyzed for establishing suitable borders of segments to obtain maximum stationary inside the segment. There are various algorithms of adaptive segmentation [20]. However, many of them aimed at the division of a single channel data. In the approach proposed in the work, was used a metric which measures similarity between multichannel data pieces. This metric is named Riemannian distance and was firstly applied in brain-computer interface (BCI) applications [21].

It utilizes covariance matrices of the data which are symmetric positive-definite (SPD) matrices and fall within the Riemannian geometry domain. Then Riemannian distance between two SPD matrices $\mathbf{P_1}$ and $\mathbf{P_2}$ in P(n) is defined as:

$$\delta_R(\mathbf{P_1}, \mathbf{P_2}) = ||Log(\mathbf{P_1^{-1}P_2})||_F = \left[\sum_{i=1}^{n} \log^2 \lambda_i\right]^{\frac{1}{2}} \tag{3.1}$$

where P(n) is a set of all $n \times n$ SPD matrices, $\mathbf{P_1}$ and $\mathbf{P_2}$ are covariance matrices of two adjacent segments, $\lambda_i$ is real positive eigenvalues of matrix $\mathbf{P_1^{-1}P_2}$, $||.||_F$ is the Frobenius norm of the matrix and Log() is logarithm of the matrix, which can be computed by diagonalization of $\mathbf{P_1^{-1}P_2}$ [21].

The proposed method involves the following steps:

1. The signal is divided into overlapped segments with constant length.
2. For every two adjacent windows is computed a covariance matrix.
3. A Riemannian distance between these two matrices is calculated. Obtained function is called DIST.
4. Borders of the segments are established in locations where function DIST calculated on the previous step reaches local peaks. Minimum peak height set at mean value of the DIST function. The minimum distance between peaks is the quotient of segment length and step.

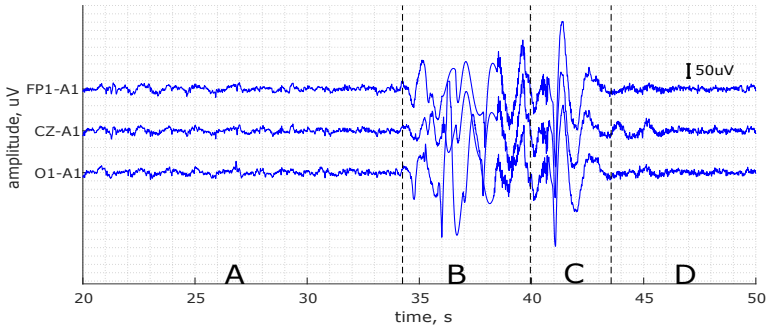
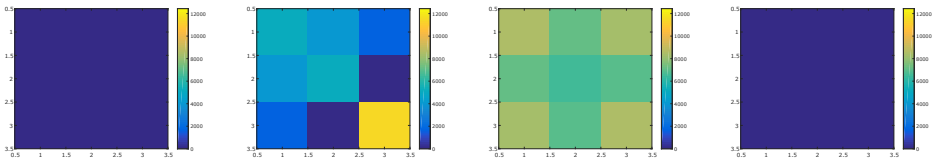



**Figure 3.3:** Result of adaptive segmentation of EEG signal. Dashed lines represent segments borders. There are presented two artefact free segments A (20-34 s) and D (44-50 s). Artefact segments are B (35-40 s) and C (41-43 s).



(a) (b) (c) (d)

**Figure 3.4:** Covariance matrices of segments presented in Figure 3.3: a - presents covariance matrix of segment A, b - presents covariance matrix of segment B, c - presents covariance matrix of segment C, d - presents covariance matrix of segment D.

In Figure 3.3 is presented part of PSG recording containing three EEG channels: FP1-A1, CZ-A1, O1-A1. Applying adaptive segmentation algorithm this part was divided into 4 segments labelled A-D in Figure 3.3. For

every segment covariance matrices was computed. They are represented in Figure 3.4. Covariance matrices B and C contain artefacts and covariance values are significantly higher than in non-artefact data. Such, covariance matrix of the A segment has values in the range [51.79; 83.78], while the matrix of the B segment's range is [20; 11435] ([6943; 8392], [24.66; 109.32] are covariance matrices' ranges of the C and the D segments respectively).

## 3.4 Feature extraction

The crucial part of the signal processing is feature extraction. Every signal can be represented as a combination of computed features [17]. These parameters represent the signal and directly affect the following classification. In Table 3.2 are presented features calculated in this approach.

| Feature | Description |
|---|---|
| Statistical features | |
| MEAN | Mean value defined as: $\mu = \frac{1}{N}\sum_{i=1}^{N} X_i$ |
| STD | Standard deviation: $\sigma_x = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} \lvert X_i - \mu\rvert^2}$ |
| MAX | Maximum value of a set |
| MIN | Minimum value of a set |
| MEDIAN | Median value of a set |
| SKEWNESS | Skewness (a measure of the asymmetry of the data around the sample mean) $s = \frac{\frac{1}{N}\sum_{i=1}^{N}(X_i-\mu)^3}{\sigma_x^3}$ |
| KURTOSIS | Kurtosis (a measure of how outlier-prone a distribution is) $s = \frac{\frac{1}{N}\sum_{i=1}^{N}(X_i-\mu)^4}{\sigma_x^4}$ |
| Band power of EEG spectra | |
| BP ALPHA | The average power in the alpha frequency range (8-13 Hz) |
| BP BETA | The average power in the beta frequency range (13-30 Hz) |
| BP DELTA | The average power in the delta frequency range (0-4 Hz) |
| BP THETA | The average power in the theta frequency range (4-8 Hz) |
| Relative band power of EEG spectra | |
| BP REL ALPHA | The percentage of the alpha power in the whole frequency interval |
| BP REL BETA | The percentage of the beta power in the whole frequency interval |
| BP REL DELTA | The percentage of the delta power in the whole frequency interval |
| BP REL THETA | The percentage of the theta power in the whole frequency interval |

**Table 3.2:** Calculated features.

## 3.5 Classification

Binary classification is the following part of signal processing, which task is classifying set into two groups. In this case, it predicts if current segment

is artefact or not. There are many algorithms, for instance, support vector machine (SVM), decision trees, neural networks etc [22]. In this approach is applied Naïve Bayes classifier.
Naïve Bayes classifier is based on applying Bayes' theorem with the strong "naive" assumption of independence between every pair of the features.

Bayes theorem:

$$P(C|x_1, ..., x_n) = \frac{P(C)P(x_1, ...x_n|C)}{P(x_1, ..., x_n)} \tag{3.2}$$

In our case, C is class, $x_1, ...x_n$ are computed features. $P(C)$ is the prior probability (probability of belonging to class without features) , $P(x_1, ..., x_n)$ is marginal probability (probability of reaching collected values by features), $P(C|x_1, ..., x_n)$ is posterior probability (the probability of belonging to class by given features), $P(x_1, ...x_n|C)$ is called likelihood (probability of reaching collected values by features when belong to class) [23].

## 3.6 Evaluation of classifier

The aim of classifier performance evaluation is to estimate the goodness of predicted labels. In this section, there are described statistical metrics for classifier performance evaluation commonly used in binary classification. There are provided two classes positive (P) and negative (N). The real labels and predicted by a classifier are compared. Four outcomes are possible. They are listed below and visually presented in Figure 3.5:

- True Positives (TP) is a number of labels correctly predicted as positive (real label is P, predicted label is P);
- True Negatives (TN) is a number of labels correctly predicted as negative (real label is N, predicted label is N);
- False Positives (FP) is a number of labels incorrectly predicted as positive (real label is N, predicted label is P);
- False Negatives (FN) is a number of labels incorrectly predicted as negative (real label is P, predicted label is N).

### 3.6.1 Statistical metrics

Recall or Sensitivity is the quotient of the number of TP and the total number of elements actually belong to the positive. High sensitivity shows the ability of classifier to detect TP cases, but says nothing about how many other items were incorrectly labelled as positive.

$$Recall = \frac{TP}{TP + FN} \tag{3.3}$$

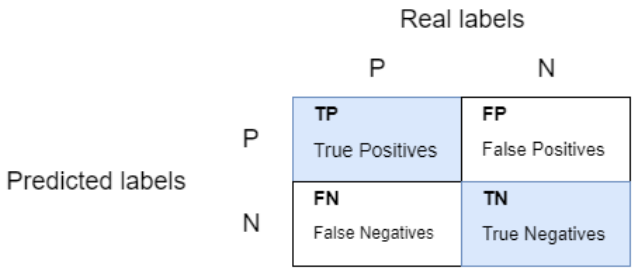



**Figure 3.5:** Confusion matrix.

Precision or Positive Predictive value is the quotient of the number of TP and the total number of elements were labelled as positive. Metric value 1.0 means that all real positive labels were predicted as positive. However, it doesn't say that anything wasn't predicted incorrectly.

$$Precision = \frac{TP}{TP + FP} \tag{3.4}$$

Specificity or True Negative Rate is the quotient of the number of TN and the total number of elements were labelled as negative. High specificity means that classifier predict less false positive labels.

$$Specificity = \frac{TN}{TN + FP} \tag{3.5}$$

Accuracy is another useful metric and it is a proportion of correctly classified (both TP and TN) among all labels.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3.6}$$

Metric Negative Predictive Value shows a quotient of true negative values and all labels marked as negative.

$$NPV = \frac{TN}{TN + FN} \tag{3.7}$$

F-measure also known as $F_1$ score is defined a harmonic mean of precision (Equation 3.4) and recall (Equation 3.3). This metric takes into account FP and FN labels and more useful than accuracy in case of unbalanced datasets [24].

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{3.8}$$

In case of artefact detection, the non-artefact class is prevalent. For this reason, in the assessment of performance classification, the conclusion is based on $F_1$ metric in this approach.

### 3.6.2 Cross-validation

After the predictive model is built, the prediction performance of the classifier may be estimated. One of the most common methods is cross-validation. The approach is based on partitioning available data into non-overlapped parts also called folds. In the work, leave-one-out-cross-validation scheme was used. One of the folds was used for model testing and rest of the folds were used to build the classifier, as a training set. Testing the build model on a testing data, statistical metrics were computed. Repeating procedure for all possible combinations of training and testing sets returns set of statistical metrics. Analysis of the metrics allows estimating the possible outcome for similar but unknown data (Figure 3.6). In the work, data of every separate subject formed a fold in order to avoid overfitting.

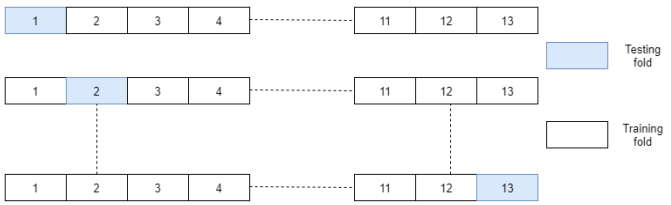
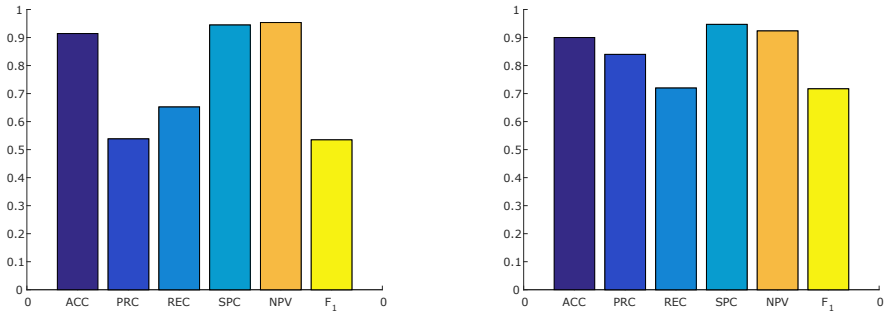



**Figure 3.6:** Leave-one-out-cross-validation scheme. Every row presents an iteration in CV. There are 13 folds in each iteration. Folds marked as white are from training set and blue ones from testing set.

# Chapter 4

# Experimental results

The proposed method was tested on the DREAM database. Leave-one-out-cross-validation scheme was applied. As classifier was chosen Naïve Bayes. Classification was provided for two types of segmentation: constant and adaptive. For evaluation of classifier performance were computed metrics such as accuracy, precision, recall, specificity, negative predictive value (NPV) and $F_1$.




(a) (b)

**Figure 4.1:** Results of classification provided (a) using constant segmentation (b) adaptive segmentation. Each bar represents average of statistical metric. From left to right accuracy (ACC), precision (PRC), recall (REC), specificity (SPC), NPV and $F_1$ respectively.

| Method | Parameters | Accuracy | Precision | Recall | Specificity | NPV | F1 |
|---|---|---|---|---|---|---|---|
| Detection procedure [7] | - | 0.89 | 0.37 | 0.64 | 0.92 | 0.96 | 0.43 |
| Constant approach | window length 3 s | 0.91 | 0.54 | 0.65 | 0.94 | 0.95 | 0.53 |
| Adaptive approach | window length 2.5 s, step 30 samples | 0.89 | 0.84 | 0.72 | 0.95 | 0.92 | 0.72 |

**Table 4.1:** Results of classification.

Results are presented in Table 4.1 and visualized in Figure 4.1. The table also contains the results of the method performed by S. Devuyst et.al.[7].

## 4.1 Constant segmentation

Classification of non-overlapped segments of length 3 seconds was performed. Evaluation of classifier performance is in Table 4.1. All the metrics are higher or similar to values provided in [7]. Probably, the authors were seeking for all possible artefacts despite the provided scoring. In the [7] number of provided detection performed in the project is much higher than in the scoring or detection provided in the study. Accuracy value is high, although $F_1$ is 0.53 what is low. This may be due to the big number of segments and the fact that dataset is unbalanced: the amount of negative labels is fifty times bigger than positive ones.
Different window size were tested: 0.3-1, 1.25, 1.5, 1.75, 2, 2.5, 3, 5, 10, 20, 30 s. The window length was chosen among the others because $F_1$ reaches the highest value. Achieved results for these lengths are presented in section A.1.

## 4.2 Adaptive segmentation

Classification of non-overlapped segments of different length was performed. Segments' borders were achieved applying constant segmentation algorithm with parameters: window length 2.5 s, step 30 samples. Evaluation of classifier performance is in Table 4.1. In this case, $F_1$ reaches 0.72. Dataset is still unbalanced, but the difference between the amount of positive and negative values is lower, what provides higher values of precision and recall. Number of segments in this classification varies depending on the subject.

For achieving best results were tested various combinations of window's lengths and steps values. Tested window lengths are 0.3-1, 1.25, 1.5, 1.75, 2, 2.5, 3 s; tested steps are in range [10; 100]. The graphic representation of achieved $F_1$ is in Figure 4.2. Detailed results of these tests are in section A.2

During the testing, there are obtained all possible outcomes and they are demonstrated in Figure 4.3. True Positive is in the segment 460-472 s. This area was labelled as artefact and was predicated as an artefact by a classifier. Segments in the beginning and in the end of the figure were correctly classified as negatives and counted as TN. False Positive outcome might be observed in segment 457-460 s. Classifier marked this part as artefact, while experts did not labelled it. Segment with borders 472 s and 480 s is an example of False Negative outcome. A detailed analysis of false outcomes is discussed below.

**False positives**

False positives outcomes may appear for different reasons. One of the examples is in Figure 4.4. In this sample segments from 194 s to 203 s are marked as false positive: predicted label is positive, but visual scoring is negative. In examining these segments, it should be acknowledged that activity in them

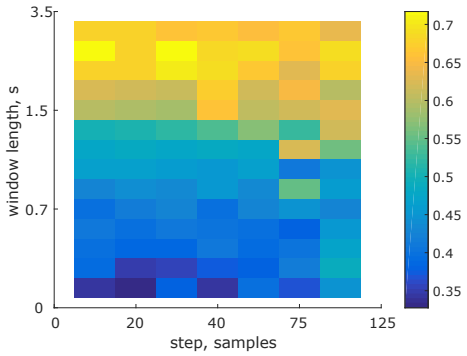



**Figure 4.2:** Graphic representation of achieved results using adaptive segmentation. Each cell display calculated $F_1$ for given window length and step. Color specifies the value of metric. Row represents window length and columns are values of step used in testing.

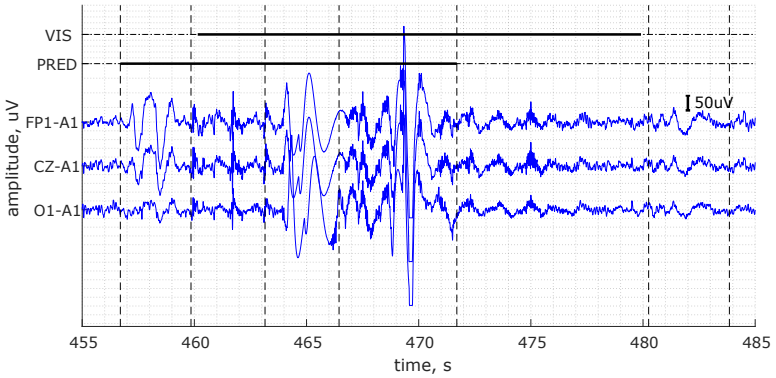


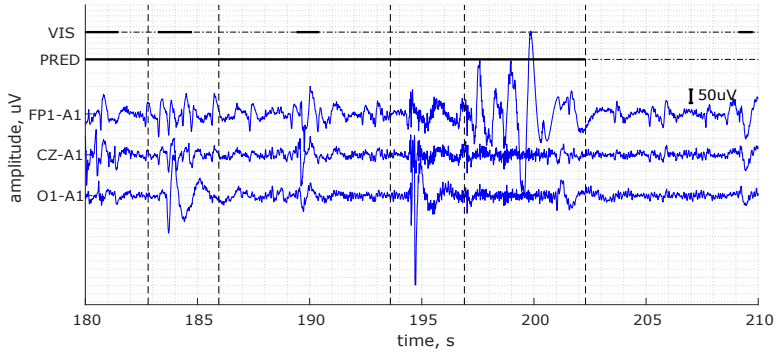

**Figure 4.3:** Example of a recording with a visual scoring by a trained expert (VIS) and prediction (PRED) by a proposed adaptive method. In a binary scoring, dashed line stands for non-artefacts and bold line for artefacts. Horizontal dashed lines present obtained adaptive segmentation.

is far from normal. However, experts hadn't marked as artefact these parts. The classification depends on labels achieved from visual scorings.

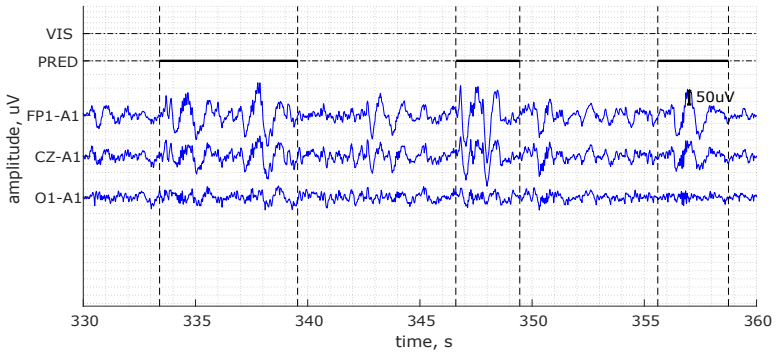
Another example of false positive outcome is in Figure 4.5. Segments are marked as damaged, a possible reason is that these segments contain k-complexes, which don't occur in training set. The probable solution is the detection of k-complexes using wavelet transformation.

**False negatives**

False negative outcomes occur more frequently than positive ones. In Figure 4.6 is sample of two small undetected artefacts in one big segment. Instead of one segment, adaptive segmentation approach was to divide this part of the recording into three: first: 48 - 51; 51- 61; 61-64,5. This is caused by minimum peak height value which is used in setting borders. This height is computed as mean value of DIST function (obtained in the 3rd step of adaptive segmentation algorithm) which are higher in the other part of the recording. Excluding this parameter will result a large number of segments.




(a)




(b)

**Figure 4.4:** Typical false positives. There are provided recordings with a visual scoring by a trained expert (VIS) and prediction (PRED) by a proposed adaptive method. In a binary scoring, dashed line stands for non-artefacts and bold line for artefacts. Horizontal dashed lines present obtained adaptive segmentation.

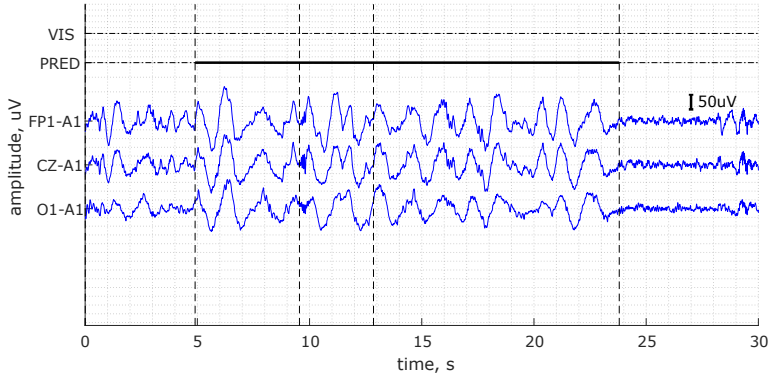



**Figure 4.5:** Typical false positive outcome. There is provided a recording with a visual scoring by a trained expert (VIS) and prediction (PRED) by a proposed adaptive method. In a binary scoring, dashed line stands for non-artefacts and bold line for artefacts. Horizontal dashed lines present obtained adaptive segmentation.

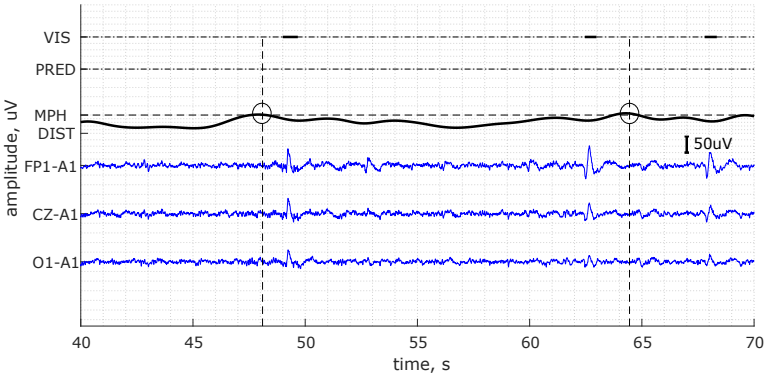



**Figure 4.6:** Typical false negative outcome. There is provided a recording with a visual scoring by a trained expert (VIS) and prediction (PRED) by a proposed adaptive method. In a binary scoring, dashed line stands for non-artefacts and bold line for artefacts. Horizontal dashed lines present obtained adaptive segmentation. DIST is a function obtained in the step 3 of algorithm. MPH is mean value of DIST function. Borders of the segments are established in locations where function DIST calculated on the previous step reaches local peaks above MPH. Obtained peaks are labeled as circles.

Classifier also has not marked as artefact segments containing activity like in Figure 4.7. Amplitude in them is close to zero. Features of the segments lies in untypical range for this training set. It is caused by the fact that these type of artefact (falling electrode) occurs only in this subject. Figure 4.8 shows that standard deviation's distribution of non-artefact segments is right-skewed. The value of this segment standard deviation is 0.36 (on the histogram is marked with asterisk). This can be fixed by adding more subjects with this type of artifact to the training set.

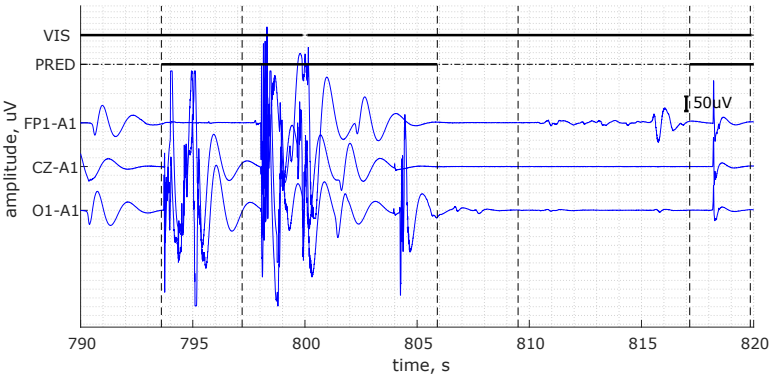



**Figure 4.7:** Typical false negative outcome. There is provided a recording with a visual scoring by a trained expert (VIS) and prediction (PRED) by a proposed adaptive method. In a binary scoring, dashed line stands for non-artefacts and bold line for artefacts. Horizontal dashed lines present obtained adaptive segmentation.

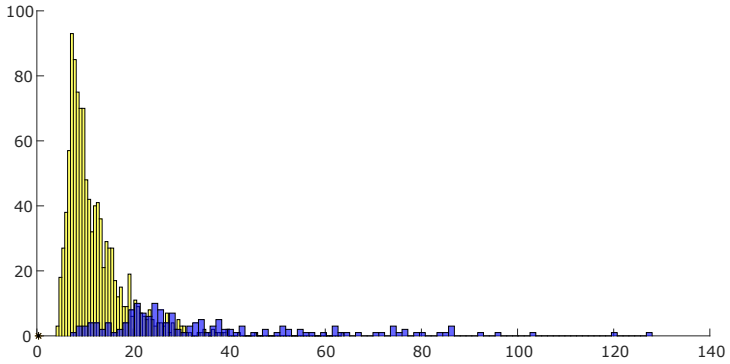



**Figure 4.8:** Histogram of standard deviation of training set. Yellow histogram is standard deviation of positive labels, blue - negative. Asterisk represents a value of standard deviation of the falling electrode artefact.

In Figure 4.9 is another example of false negative outcome: segment from 280 to 291 s isn't pointed as positive. This can be explained as follows: the amplitude of activity in this segment is not that high as in comparison with adjacent ones.

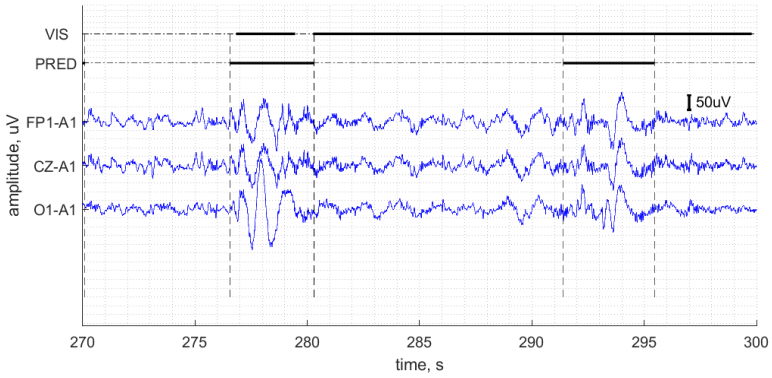
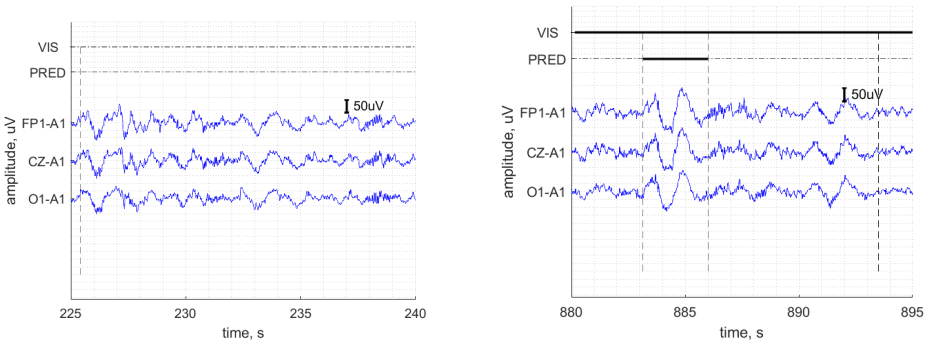



**Figure 4.9:** Typical false negative outcome. There is provided a recording with a visual scoring by a trained expert (VIS) and prediction (PRED) by a proposed adaptive method. In a binary scoring, dashed line stands for non-artefacts and bold line for artefacts. Horizontal dashed lines present obtained adaptive segmentation.

The dependence of the classification on excerpts scoring contributes to the occurrence not only of false positives, but also false negatives outcomes. The example of this effect is in Figure 4.10. The activity looks similar on both samples, but on (a) it is not counted as artefact by experts.




(a) (b)

**Figure 4.10:** Comparison of normal activity marked as non-artefact (a) and artefact (b). There is provided a recording with a visual scoring by a trained expert (VIS) and prediction (PRED) by a proposed adaptive method. In a binary scoring, dashed line stands for non-artefacts and bold line for artefacts. Horizontal dashed lines present obtained adaptive segmentation.

# Chapter 5

# Discussion and Conclusion

All the goals of this thesis were accomplished. A new method of artefact detection was developed. This approach is based on classification on data segments obtained with new proposed multichannel adaptive segmentation. Produced segment borders are placed where the difference in a signal to the left and to the right is maximal in local terms. This was accomplished using Riemannian distance, a method popular in a BCI field. Classification was provided with Naïve Bayes classifier. Testing were performed on 13 subjects from open source The DREAMS Artifacts Database.

Various statistical metrics were calculated during leave-one-out-cross-validation procedure. The results of developed method were compared with results of a common method based on constant segmentation and other research performed by S. Devuyst. There was shown a significant increase in comparison with other methods. Utilizing of state of art method has brought reliable results: $F_1$ is 53 %. The method based on applying specific algorithms for each type of artefact for the EEG channels used in this work has $F_1$ equal to 43 %. The elaborated method's $F_1$ reaches 72 %. That was achieved due to concentration of similar statistical properties inside a single segment.

However, there were found some challenges during the testing session. Data used in the study contains different sleep stages and statistical properties. It, consequently, leads to the difficulty of applying the same segmentation parameters to all subjects in order to get better results. Besides, the results were often spoiled by the occurrence of delta waves with large amplitude. Possible solution of the problem is filtration of the signal before segmentation but signals should be restored for clinical purposes. To determine the sleep stage N3 or k-complexes in N2, delta waves are highly important. Moreover, visual scoring provided by an expert was confusing sometimes. There were activity visually similar but in one case labeled as an artefact and ignored in other one. Those cases occurred rarely though.

In future, other methods of automatic classification could be applied, for instance, SVM and neural networks. Clustering (unsupervised learning) is another approach that can be implemented for artefact detection. Also, a new method is possible to test on other types of artefacts or just on one; or on other channels. For example, EOG channels. After achieving higher results the algorithm can be extended to real-time detection.

# Bibliography


[1] S. Sanei and J. Chambers, *EEG signal processing*. Hoboken, NJ: John Wiley & Sons, c2007.

[2] E. Garcia-Rill, "Neural mechanisms of sleep and circadian rhythms," 06 2012.

[3] J. R. D. Robert L. Wilkins, James R. Dexter Robert L. Wilkins and A. J. H. consulting editor, *Clinical assessment in respiratory care*. St. Louis, Mo: Mosby, 6th ed ed., 2010.

[4] S. Motamedi-Fakhr, M. Moshrefi-Torbati, M. Hill, C. M. Hill, and P. R. White, "Signal processing techniques applied to human sleep eeg signals—a review," *Biomedical Signal Processing and Control*, vol. 10, pp. 21–33, 2014.

[5] M. K. Islam, A. Rastegarnia, and Z. Yang, "Methods for artifact detection and removal from scalp eeg," *Neurophysiologie Clinique/Clinical Neurophysiology*, vol. 46, no. 4-5, pp. 287–305, 2016.

[6] M. Roháľová, P. Sykacek, M. Koska, and G. Dorffner, "Detection of the eeg artifacts by the means of the (extended) kalman filter," 2001.

[7] S. Devuyst, T. Dutioit, T. Ravet, and P. Stenuit, "Automatic processing of eeg-eog-emg artifacts in sleep stage classification," 2008.

[8] A. Hyvärinen and E. Oja, "Independent component analysis," *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[9] H. Nolan, R. Whelan, and R. Reilly, "Faster," *Journal of Neuroscience Methods*, vol. 192, no. 1, pp. 152–162, 2010.

[10] S. O'Regan, S. Faul, and W. Marnane, "Automatic detection of eeg artefacts arising from head movements using eeg and gyroscope signals," *Medical Engineering & Physics*, vol. 35, no. 7, pp. 867–874, 2013.

[11] L. Trans Cranial Technologies, "10/20 system positioning manual," 2012.

[12] B. V. Vaughn, "Technical review of polysomnography," *CHEST Journal*, vol. 134, no. 6, pp. 1310–, 2008-12-01.

[13] R. B. Berry, R. Brooks, C. Gamaldo, S. M. Harding, R. M. Lloyd, S. F. Quan, M. T. Troester, and B. V. Vaughn, "Aasm scoring manual updates for 2017 (version 2.4)," *Journal of Clinical Sleep Medicine*, vol. 13, no. 05, pp. 665–666, 2017-5-15.

[14] R. Sucholeiki, "Normal eeg waveforms: Overview, frequency, morphology," 2017.

[15] S. R. Benbadis, "Normal sleep eeg: Overview, stage i sleep, stage ii sleep," 2018.

[16] J. C. Segen, *The dictionary of modern medicine*. Parthenon, 1992.

[17] V. Gerla, "Automated analysis of long-term eeg signals," 2012.

[18] H. Abdullah, N. C. Maddage, I. Cosic, and D. Cvetkovic, "Cross-correlation of eeg frequency bands and heart rate variability for sleep apnoea classification," *Medical & Biological Engineering & Computing*, vol. 48, pp. 1261–1269, Dec 2010.

[19] M. O. Mendez, I. Chouvarda, A. Alba, A. M. Bianchi, A. Grassi, E. Arce-Santana, G. Milioli, M. G. Terzano, and L. Parrino, "Analysis of a-phase transitions during the cyclic alternating pattern under normal sleep," *Medical & Biological Engineering & Computing*, vol. 54, pp. 133–148, Jan 2016.

[20] J. Barlow, O. Creutzfeldt, D. Michael, J. Houchin, and H. Epelbaum, "Automatic adaptive segmentation of clinical eegs," *Electroencephalography and Clinical Neurophysiology*, vol. 51, no. 5, pp. 512–525, 1981.

[21] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Riemannian geometry applied to BCI classification," in *9th International Conference Latent Variable Analysis and Signal Separation (LVA/ICA 2010)* (V. M. E. G. R. V. E. E. Vigneron, V.; Zarzoso, ed.), vol. 6365 of *Series: Lecture Notes in Computer Science. Subseries: Theoretical Computer Science and General Issues*, (Saint-Malo, France), pp. 629–636, Springer, Sept. 2010. ISBN 978-3-642-15994-7, Softcover.

[22] R. Boostani, F. Karimzadeh, and M. Nami, "A comparative review on sleep stage classification methods in patients and healthy individuals," *Computer Methods and Programs in Biomedicine*, vol. 140, pp. 77–91, 2017.

[23] L. Wasserman, *All of statistics*. New York: Springer, corrected second print ed., 2010.

[24] D. Powers, "Evaluation: From precision, recall and f-factor to roc, informedness, markedness correlation," vol. 2, 01 2008.

# Appendices

# Appendix A

# Tables of results

## A.1 Constant segmentation

| Window length | Accuracy | Precision | Recall | Specificity | NPV | $F_1$ |
|---|---|---|---|---|---|---|
| 0.30 | 0.91 | 0.58 | 0.49 | 0.96 | 0.94 | 0.47 |
| 0.40 | 0.91 | 0.59 | 0.51 | 0.96 | 0.94 | 0.49 |
| 0.50 | 0.91 | 0.57 | 0.52 | 0.95 | 0.94 | 0.48 |
| 0.60 | 0.91 | 0.57 | 0.53 | 0.95 | 0.95 | 0.49 |
| 0.70 | 0.91 | 0.57 | 0.55 | 0.95 | 0.95 | 0.50 |
| 0.80 | 0.91 | 0.57 | 0.60 | 0.95 | 0.95 | 0.51 |
| 0.90 | 0.91 | 0.59 | 0.60 | 0.95 | 0.95 | 0.53 |
| 1.00 | 0.91 | 0.58 | 0.56 | 0.95 | 0.95 | 0.51 |
| 1.25 | 0.92 | 0.61 | 0.58 | 0.95 | 0.95 | 0.52 |
| 1.50 | 0.91 | 0.53 | 0.63 | 0.94 | 0.95 | 0.48 |
| 1.75 | 0.91 | 0.54 | 0.60 | 0.95 | 0.95 | 0.49 |
| 2.00 | 0.92 | 0.53 | 0.59 | 0.95 | 0.95 | 0.49 |
| 2.50 | 0.91 | 0.54 | 0.65 | 0.94 | 0.95 | 0.52 |
| **3.00** | **0.91** | **0.54** | **0.65** | **0.94** | **0.95** | **0.53** |
| 5.00 | 0.91 | 0.37 | 0.47 | 0.94 | 0.96 | 0.37 |
| 10.00 | 0.91 | 0.27 | 0.57 | 0.93 | 0.97 | 0.34 |
| 20.00 | 0.89 | 0.20 | 0.43 | 0.91 | 0.97 | 0.24 |
| 30.00 | 0.91 | 0.15 | 0.25 | 0.93 | 0.97 | 0.15 |

**Table A.1:** Evaluation of classification with constant segmentation for different window lengths

## A.2 Adaptive segmentation

**Table A.2:** Evaluation of classification with adaptive segmentation for different window lengths

| Win. length | Step | Accuracy | Precision | Recall | Specificity | NPV | $F_1$ |
|---|---|---|---|---|---|---|---|
| 0.30 | 10.00 | 0.90 | 0.34 | 0.46 | 0.94 | 0.95 | 0.34 |
| | 20.00 | 0.90 | 0.30 | 0.48 | 0.93 | 0.95 | 0.33 |
| | 30.00 | 0.90 | 0.34 | 0.56 | 0.93 | 0.95 | 0.38 |
| | 40.00 | 0.90 | 0.34 | 0.47 | 0.93 | 0.95 | 0.34 |
| | 50.00 | 0.90 | 0.41 | 0.52 | 0.93 | 0.95 | 0.40 |
| | 75.00 | 0.90 | 0.31 | 0.67 | 0.93 | 0.95 | 0.37 |
| | 100.00 | 0.90 | 0.44 | 0.60 | 0.93 | 0.95 | 0.44 |
| 0.40 | 10.00 | 0.90 | 0.39 | 0.53 | 0.93 | 0.95 | 0.39 |
| | 20.00 | 0.90 | 0.32 | 0.49 | 0.94 | 0.95 | 0.35 |
| | 30.00 | 0.90 | 0.31 | 0.51 | 0.93 | 0.95 | 0.35 |
| | 40.00 | 0.90 | 0.35 | 0.50 | 0.93 | 0.95 | 0.37 |
| | 50.00 | 0.90 | 0.36 | 0.51 | 0.93 | 0.94 | 0.38 |
| | 75.00 | 0.90 | 0.40 | 0.52 | 0.94 | 0.95 | 0.42 |
| | 100.00 | 0.90 | 0.47 | 0.63 | 0.93 | 0.95 | 0.49 |
| 0.50 | 10.00 | 0.90 | 0.41 | 0.48 | 0.94 | 0.94 | 0.40 |
| | 20.00 | 0.89 | 0.38 | 0.49 | 0.93 | 0.94 | 0.39 |
| | 30.00 | 0.90 | 0.38 | 0.48 | 0.94 | 0.94 | 0.38 |
| | 40.00 | 0.90 | 0.43 | 0.49 | 0.93 | 0.94 | 0.41 |
| | 50.00 | 0.90 | 0.39 | 0.49 | 0.93 | 0.94 | 0.39 |
| | 75.00 | 0.89 | 0.40 | 0.51 | 0.93 | 0.94 | 0.40 |
| | 100.00 | 0.90 | 0.46 | 0.58 | 0.93 | 0.95 | 0.47 |
| 0.60 | 10.00 | 0.90 | 0.41 | 0.50 | 0.93 | 0.94 | 0.41 |
| | 20.00 | 0.89 | 0.39 | 0.49 | 0.93 | 0.94 | 0.40 |
| | 30.00 | 0.89 | 0.40 | 0.51 | 0.93 | 0.94 | 0.41 |
| | 40.00 | 0.89 | 0.39 | 0.51 | 0.93 | 0.94 | 0.40 |
| | 50.00 | 0.89 | 0.42 | 0.48 | 0.93 | 0.93 | 0.40 |
| | 75.00 | 0.89 | 0.39 | 0.46 | 0.93 | 0.93 | 0.38 |
| | 100.00 | 0.90 | 0.48 | 0.54 | 0.93 | 0.94 | 0.45 |
| 0.70 | 10.00 | 0.89 | 0.40 | 0.48 | 0.93 | 0.93 | 0.40 |
| | 20.00 | 0.89 | 0.42 | 0.50 | 0.93 | 0.93 | 0.42 |
| | 30.00 | 0.89 | 0.45 | 0.51 | 0.93 | 0.93 | 0.43 |
| | 40.00 | 0.88 | 0.42 | 0.49 | 0.92 | 0.93 | 0.40 |
| | 50.00 | 0.89 | 0.46 | 0.51 | 0.93 | 0.93 | 0.43 |
| | 75.00 | 0.88 | 0.45 | 0.61 | 0.93 | 0.93 | 0.45 |
| | 100.00 | 0.89 | 0.44 | 0.54 | 0.93 | 0.94 | 0.43 |
| 0.80 | 10.00 | 0.89 | 0.46 | 0.52 | 0.93 | 0.93 | 0.43 |
| | 20.00 | 0.89 | 0.50 | 0.53 | 0.93 | 0.93 | 0.44 |
| | 30.00 | 0.88 | 0.50 | 0.50 | 0.93 | 0.93 | 0.44 |
| | 40.00 | 0.89 | 0.49 | 0.54 | 0.93 | 0.93 | 0.45 |
| | 50.00 | 0.88 | 0.50 | 0.52 | 0.93 | 0.93 | 0.44 |
| | 75.00 | 0.89 | 0.57 | 0.64 | 0.93 | 0.93 | 0.55 |
| | 100.00 | 0.89 | 0.48 | 0.54 | 0.93 | 0.93 | 0.46 |
| 0.90 | 10.00 | 0.89 | 0.50 | 0.55 | 0.93 | 0.93 | 0.46 |
| | 20.00 | 0.89 | 0.50 | 0.55 | 0.93 | 0.93 | 0.47 |
| | 30.00 | 0.89 | 0.49 | 0.54 | 0.93 | 0.93 | 0.46 |
| | 40.00 | 0.89 | 0.52 | 0.52 | 0.94 | 0.92 | 0.45 |
| | 50.00 | 0.89 | 0.51 | 0.52 | 0.93 | 0.93 | 0.46 |
| | 75.00 | 0.88 | 0.46 | 0.50 | 0.93 | 0.92 | 0.41 |

**Table A.2 – continued from previous page**

| Win. length | Step | Accuracy | Precision | Recall | Specificity | NPV | $F_1$ |
|---|---|---|---|---|---|---|---|
| | 100.00 | 0.88 | 0.48 | 0.54 | 0.93 | 0.92 | 0.44 |
| 1.00 | 10.00 | 0.88 | 0.52 | 0.55 | 0.93 | 0.92 | 0.48 |
| | 20.00 | 0.89 | 0.53 | 0.53 | 0.93 | 0.92 | 0.48 |
| | 30.00 | 0.89 | 0.52 | 0.53 | 0.93 | 0.92 | 0.48 |
| | 40.00 | 0.89 | 0.52 | 0.54 | 0.94 | 0.93 | 0.48 |
| | 50.00 | 0.89 | 0.52 | 0.53 | 0.93 | 0.92 | 0.47 |
| | 75.00 | 0.89 | 0.64 | 0.73 | 0.93 | 0.93 | 0.62 |
| | 100.00 | 0.89 | 0.60 | 0.67 | 0.94 | 0.92 | 0.56 |
| 1.25 | 10.00 | 0.89 | 0.60 | 0.54 | 0.94 | 0.92 | 0.50 |
| | 20.00 | 0.89 | 0.60 | 0.54 | 0.94 | 0.92 | 0.51 |
| | 30.00 | 0.89 | 0.60 | 0.55 | 0.93 | 0.92 | 0.52 |
| | 40.00 | 0.89 | 0.57 | 0.66 | 0.93 | 0.93 | 0.54 |
| | 50.00 | 0.89 | 0.63 | 0.67 | 0.93 | 0.92 | 0.57 |
| | 75.00 | 0.89 | 0.57 | 0.66 | 0.93 | 0.93 | 0.52 |
| | 100.00 | 0.89 | 0.66 | 0.75 | 0.93 | 0.93 | 0.62 |
| 1.50 | 10.00 | 0.89 | 0.67 | 0.70 | 0.94 | 0.93 | 0.60 |
| | 20.00 | 0.89 | 0.68 | 0.70 | 0.94 | 0.92 | 0.59 |
| | 30.00 | 0.89 | 0.65 | 0.70 | 0.94 | 0.93 | 0.58 |
| | 40.00 | 0.90 | 0.72 | 0.74 | 0.94 | 0.93 | 0.66 |
| | 50.00 | 0.89 | 0.64 | 0.75 | 0.94 | 0.93 | 0.61 |
| | 75.00 | 0.90 | 0.66 | 0.75 | 0.94 | 0.93 | 0.62 |
| | 100.00 | 0.89 | 0.71 | 0.72 | 0.94 | 0.93 | 0.63 |
| 1.75 | 10.00 | 0.89 | 0.68 | 0.71 | 0.94 | 0.92 | 0.63 |
| | 20.00 | 0.89 | 0.67 | 0.70 | 0.93 | 0.92 | 0.62 |
| | 30.00 | 0.89 | 0.65 | 0.71 | 0.94 | 0.92 | 0.61 |
| | 40.00 | 0.90 | 0.75 | 0.73 | 0.94 | 0.93 | 0.67 |
| | 50.00 | 0.89 | 0.65 | 0.75 | 0.94 | 0.93 | 0.62 |
| | 75.00 | 0.89 | 0.70 | 0.72 | 0.94 | 0.92 | 0.65 |
| | 100.00 | 0.89 | 0.65 | 0.71 | 0.94 | 0.92 | 0.60 |
| 2.00 | 10.00 | 0.90 | 0.78 | 0.69 | 0.95 | 0.92 | 0.68 |
| | 20.00 | 0.90 | 0.75 | 0.71 | 0.95 | 0.93 | 0.68 |
| | 30.00 | 0.90 | 0.81 | 0.71 | 0.95 | 0.93 | 0.70 |
| | 40.00 | 0.90 | 0.75 | 0.74 | 0.94 | 0.93 | 0.69 |
| | 50.00 | 0.91 | 0.77 | 0.69 | 0.95 | 0.93 | 0.67 |
| | 75.00 | 0.89 | 0.73 | 0.67 | 0.94 | 0.92 | 0.63 |
| | 100.00 | 0.90 | 0.76 | 0.71 | 0.95 | 0.92 | 0.68 |
| 2.50 | 10.00 | 0.90 | 0.83 | 0.72 | 0.94 | 0.92 | 0.71 |
| | 20.00 | 0.89 | 0.77 | 0.71 | 0.94 | 0.92 | 0.68 |
| | 30.00 | 0.90 | 0.84 | 0.72 | 0.95 | 0.92 | 0.72 |
| | 40.00 | 0.89 | 0.78 | 0.71 | 0.94 | 0.92 | 0.68 |
| | 50.00 | 0.90 | 0.79 | 0.72 | 0.94 | 0.92 | 0.69 |
| | 75.00 | 0.89 | 0.76 | 0.70 | 0.94 | 0.92 | 0.66 |
| | 100.00 | 0.90 | 0.79 | 0.71 | 0.94 | 0.93 | 0.69 |
| 3.00 | 10.00 | 0.88 | 0.79 | 0.70 | 0.93 | 0.91 | 0.68 |
| | 20.00 | 0.88 | 0.79 | 0.70 | 0.93 | 0.91 | 0.68 |
| | 30.00 | 0.88 | 0.77 | 0.68 | 0.93 | 0.91 | 0.66 |
| | 40.00 | 0.89 | 0.79 | 0.68 | 0.94 | 0.92 | 0.67 |
| | 50.00 | 0.88 | 0.75 | 0.68 | 0.93 | 0.92 | 0.65 |
| | 75.00 | 0.89 | 0.75 | 0.72 | 0.94 | 0.92 | 0.67 |
| | 100.00 | 0.89 | 0.73 | 0.70 | 0.94 | 0.92 | 0.64 |
| End of Table | | | | | | | |

# Appendix B

# Content of the CD

CD contains three folders:

- data - data with recordings and scorings; variables needed for classification.
- functions - implemented algorithms.
- covariancetoolbox-master - covariance toolbox for Matlab, including Riemannian geometry implemented by Alexandre Barachant

For running classification add to path all folders and run test_class.m in Matlab. It is also required covariance toolbox to be activated.