Diplomarbeit

Automatic detection of sleep arousals and investigation on the relation with leg movements

Ausgeführt am Institut für

Stochastik und Wirtschaftsmathematik

der Technischen Universität Wien

unter der Anleitung von

Ao. Univ. Prof. Wolfgang Scherrer

durch

Stefanie Schreiner, BSc

Wurmbrandgasse 8/2/2
1220 Wien

Wien, April 2017 Stefanie Schreiner

Acknowledgements

For the development of this thesis I would like to express my deep gratitude to Ao. Univ. Prof. Wolfgang Scherrer, who supervised it and helped me during the process. I would also like to thank Heinrich Garn, who gave me the opportunity to write this thesis during a research activity at the *AIT Austrian Institute of Technology* and Bernhard Kohn for his support regarding implementation issues.

I also want to acknowledge all my colleagues, who have accompanied me during all these years and formed a great team that helped me to accomplish many goals.

Finally, I would like to thank everyone who contributed to this thesis and especially my family for their unconditional support during the years of study.

Abstract

This thesis provides a method for the automatic detection of arousals in EEG-signals as well as an analysis of the possible relation between leg movements and arousals. For this purpose, seven patients of AKH Wien are considered: three without diagnosis and four with Periodic Leg Movement Syndrome.

The detection algorithm consists of two main steps: Firstly, the recording is divided in segments of three seconds and these are classified with a Support Vector Machine (SVM) as arousal start or no-start segments. The features extracted for each segment are mainly frequency based according to the definition of an arousal of the *AASM American Academy of Sleep Medicine*. In addition, newly investigated feature sets, based on AR-models and statistical tests, are implemented. Secondly, the exact position and length of the detections are computed and further arousal criteria are checked.

For the validation of the classifier for the arousal start segments a Leave-One-Out-Cross validation is performed and parameters for the SVM are chosen out of a grid according to the Youden-Index of the average performance for the seven patients. On average a sensitivity and specificity of 96% (classification performance of start segment detection on segments of three seconds) and a positive predictive value of 24% are reached. After the second step of the algorithm the number of False Positives is reduced and the detection of the definite final arousals performed on average with a sensitivity of 86% and a positive predictive value of about 60%.

Moreover, an analysis on the relation between arousals and leg movements is done. It can be supported that events are dependent from each other, but that the relation follows a complex mechanism rather than a simple causality. In addition, the relation between the intensity of a leg movement and the occurrence of an associated arousal is investigated. As a measure for intensity, the duration of a leg movement and a value computed from the 3D detection of leg movements, which can be interpreted as the size of a leg movement, are tested. An interesting finding is, that with both intensity values it is shown that more intense leg movements are more likely to occur with arousals. From a clinical point of view a leg movement is more critical when it causes more disruptions of sleep and an associated arousal is considered as a kind of wakefulness, which could reduce the quality of sleep. The performance of the 3D detection is supported, because of the finding that the number of leg movements associated with an arousal is about 50% less for not 3D detected leg movements than for the ones that were 3D detected. Hence, it can be supposed that most of the not 3D detected leg movements are less relevant for the diagnosis of sleep disorders.

Content

1 Introduction

1.1 Motivation

Sleep plays an important role in everybody's health throughout the whole life. The quality of sleep influences mental and physical condition and can be considered as one of the key points for our well-being. The famous sleep expert Allan Rechtschaffen said in 1971:

"If sleep doesn`t serve some vital function, it is the biggest mistake evolution ever made"

The relevance of sleep medicine is steadily increasing as the consciousness of the importance of sleep for our health is raising. Nowadays, more and more people suffer from several sleep disorders and/or problems with falling or staying asleep. Possible reasons for these disorders can be analyzed in sleep laboratories where the sleep of the patient is monitored and a wide range of bio signals is recorded.

In this thesis, the detection of so called "arousals" and their relation to leg movements will be investigated. Arousals belong to microstructural patterns in the EEG-signal that characterize a kind of wakefulness during sleep but rarely result in actual awakening. Nevertheless, the quality of sleep can be reduced, which can result in daytime fatigue and somnolence as well as less productivity and further diseases. Here the definition of the American Academy of Sleep Medicine (AASM) will be used: "Score arousals […] if there is an abrupt shift of EEG frequency including alpha, theta and/or frequencies greater than 16 Hz (but not spindles) that lasts at least 3 seconds, with at least 10 seconds of stable sleep preceding the change. […]" (Berry, et al., 2016)

Even though this definition is widely used by sleep experts for visual scoring it turns out not to be very precise when it comes to the automatic analyzation of EEG signal and leaves space for interpretation. Therefore, a reproducible and accurate automation of this process is highly desirable and would help to overcome manual scoring issues.

One of the common sleep disorders is the Periodic Limb Movement Disorder (PLMD), in which the patient's sleep is affected by frequent and periodic leg or hand movements. Abnormalities like periodic leg movement are only diagnosed as a sleep disorder if the patient suffers from actual sleep problems, like for example a high number of arousals and consequential daytime sleepiness. For this reason, the relation between leg movements and arousals is of high interest and will be investigated in the second part of this work.

1.2 Structure of the thesis

The work is generally divided in two main parts: *Materials and Methods* and *Results.* Within the first part the used algorithms and methods are derived. The results are presented in the second part of the work.

More precisely, the chapter *Materials and Methods* is further divided in the explanation of the used data set, the methods for the arousal detection and the methods for the investigation of the relation between leg movements and arousals.

The section *Methods for arousal detection* provides detailed explanations about signal preprocessing and filtering, feature extraction and analyzation, classification of arousal start segments as well as the process of the final determination of the detected arousals. Moreover, the section *Methods for the investigation on the relation between arousals and leg movements* is concerned with the analysis of the general possible dependence of the two events, a cause/effect relation between them and the relation between the intensity of a leg movement and its association with an arousal.

The results of all sections are provided with the same structure in the chapter *Results.*

For better readability, numbers will not be written in full word form and the term “seconds” will be shortened to “s”.

1.3 Polysomnographic terms and definitions

Polysomnography (PSG)

Polysomnography is a type of sleep study, usually consisting in recording several bio signals summarized in the polysomnogram in sleep laboratories to diagnose different sleep disorders. Commonly used signals are different body functions like brain activity (Electroencephalogram), eye movements, muscle activity (Electromyogram), heart rate or respiratory signals. Several tests are performed while the patient is sleeping and a whole night recording describing his body functions is stored. (Chaudhary, 2007)

Electroencephalography (EEG)

Electroencephalography is a method to record electrical brain activity. Although the EEG alone is not sufficient for doing an overall sleep analysis it still plays a key role in the polysomnogram. It’s done by placing several electrodes along the scalp and measuring voltage fluctuation over time. The analysis of this signal is done by looking at its spectral content and filtering different frequency bands (“brain waves”) out of the signal for further diagnosis. (Chaudhary, 2007)

There are different methods to place the electrodes along the scalp but one of the most common and the used one here is the international 10-20 system:

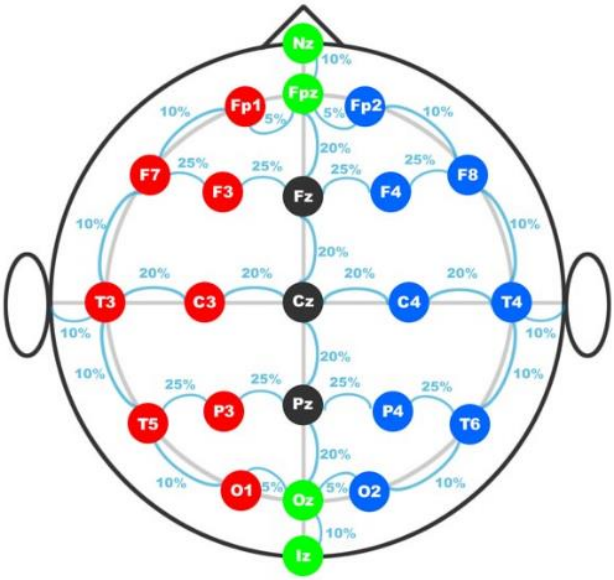


Figure 1: Placement of electrodes along the scalp with respect to the 10-20 system. (Trans Cranial Technologies ldt., 2016)

Electrocardiography (ECG)

Electrocardiography is a method to record the electrical activity of the heart. Electrodes are placed on the patient's limbs and on his chest to record electrical changes on the skin that occur from the heart muscle.

Electromyography (EMG)

Electromyography is a method to record the electrical activity that arise from skeletal muscles. It is an important tool for analyzing muscle activity to diagnose disorders of motor control. In sleep medicine, it is important for the analysis of leg (or arm) movements during sleep. The ***periodic limb movement (PLM)*** of sleep (PLMS) describes the frequent occurrence of involuntary movement of legs (or arms) during sleep. If it leads to daytime sleepiness or problems with staying or falling asleep, the patient is diagnosed with periodic limb movement disorder (PLMD).

Leg Movement (LM) and PLM series

The scoring of a LM and further of a series of consecutive LM events is defined by the AASM as follows:

- *The following defines a significant leg movement (LM) event:*
    a. *The minimum duration of a LM event is 0.5 seconds.*
    b. *The maximum duration of a LM event is 10 seconds.*
    c. *The minimum amplitude of a LM event is an 8 µV increase in EMG voltage above resting EMG (duration of at least 0.5 seconds).*
    d. *The timing of the onset of a LM event is defined as the point at which there is an 8 µV increase in EMG voltage above resting EMG.*

e. *The timing of the ending of a LM event is defined as the start of a period lasting at least 0.5 seconds during which the EMG does not exceed 2 µV above resting EMG.*

- *The following defines a PLM series:*
  - *The minimum number of consecutive LM events needed to define a PLM series is 4 LMs.*
  - *The period length between LMs (defined as the time between onsets of consecutive LMs) to include them as part of a PLM series is 5 to 90 seconds.*
  - *Leg movements on 2 different legs separated by less than 5 seconds between movement onsets are counted as a single leg movement. The period length to the next LM following this group of LMs is measured from the onset of the first LM to the onset of the next.*

(Berry, et al., 2016)

Hypnogram and characteristic frequency bands in sleep medicine

Sleep consists of nonrapid eye movement (NREM) and rapid eye movement (REM) sleep. NREM sleep is divided in stages 1-4 and REM sleep is also often called sleep stage 5. The sleep stages are visualized in the hypnogram, where the number of the current stage is displayed as a constant, changing with sleep stages. Stages 1 and 2 are characteristic for light sleep and stage 3 and 4 are called deep or slow-wave sleep. Sometimes stage 3 and 4 are considered as one sleep stage (for example in the American Academy of Sleep Medicine). Periods of awakening may also interrupt sleep during the night and are marked in the hypnogram with zero.

Sleep stage scoring is done by sleep experts visualizing successively 30s-epochs of EEG signal and deciding to which sleep stage most these 30s are belonging to. In some institutions there already exists a software for automatical staging that also considers 30s-epochs. The American Academy of Sleep Medicine (AASM) released a variety of rules for sleep staging. (Carney MD, Geyer MD, & Sachin PHD, 2012) A major throwback of these rules is that they were developed for manual analyzing in 30s epochs and therefore abrupt changes in the hypnogram could mean that the real change in sleep stage could have happened at any time in the last epoch.

In sleep medicine when working with EEG data the following characteristic frequency bands – according to the AASM rules - are analyzed (Berry, et al., 2016):

- *Delta δ (0 – 3.99 Hz): typical for dreamless deep sleep*
- *Theta θ (4 – 7.99 Hz): typical for light sleeping stage*
- *Alpha α (8 – 13 Hz): typical for being awake but with closed eyes*
- *Sigma σ (11 – 16Hz): typical for the so called sleep spindles*
- *Beta β (13 – 30 Hz): typical for being awake with eyes open but no special concentration*

For the implementation, the following slightly adjusted bands will be used:

- *Delta δ (0.4 – 4 Hz): typical for dreamless deep sleep*
- *Theta θ (4 – 8 Hz): typical for light sleeping stage*
- *Alpha α (8 – 12 Hz): typical for being awake but with closed eyes*
- *Sigma σ (12 – 16 Hz): typical for so called sleep spindles*

- *Beta β (16 – 30 Hz): typical for being awake with eyes open but no special concentration*

Further the whole signal is filtered within a frequency band of 0.4 to 40 Hz because frequencies greater than 40 Hz are not commonly analyzed frequencies in sleep analysis.

Sleep spindle

A sleep spindle (also called sigma waves) is a burst or change of EEG signal that occurs during sleep stage 2 and is often seen over central or frontal head regions. It consists of 11 – 16 Hz (mostly 12 – 14 Hz) waves that occur for at least 0.5 seconds. Sleep spindles represent brain activity that intents to keep the patient in a tranquil condition. (Berry, et al., 2016)

K-Komplex

A K-Komplex is a negative wave with a sharp increase until it reaches a maximum of amplitude and a directly following decrease and a return to the zero line. The duration should be at least 0.5 seconds and it usually occurs in sleep stage 2 and can be seen over central head regions. The maximum amplitude that is reached is normally about 75 µV but can also reach values above 200 µV. (Berry, et al., 2016)

Arousal

In literature one can find various definitions of an arousal. In general, an arousal is a kind of wakefulness during sleep but rarely resulting in actual awakening. They are rather like K-Komplexes and sleep spindles kind of patterns found in non-stationary EEG epochs, mostly found in sleep stage 2 or REM sleep. (De Carli, Nobili, Gelcich, & Ferrilo, 1999) Nevertheless, the quality of sleep can be reduced, which can result in daytime fatigue and somnolence as well as less productivity and further diseases. In this work the widely common definition of the American Academy of Sleep Medicine will be used:

"Score arousals during sleep stages N1, N2, N3 or R[1] if there is an abrupt shift of EEG frequency including alpha, theta and/or frequencies greater than 16 Hz (but not spindles) that lasts at least 3 seconds, with at least 10 seconds of stable sleep preceding the change. Scoring of arousal during REM requires a concurrent increase in submental EMG lasting at least 1 second.

Note 1. Arousal scoring should incorporate information from the frontal, central, and occipital derivations.

Note 2. Arousal scoring can be improved by the use of additional information in the recording such as respiratory events and/or additional EEG channels. Scoring of arousals, however, cannot be based on this additional information alone and such information does no modify any of the arousal scoring rules.

Note 3. Arousals meeting all scoring criteria but occurring during an awake epoch in the recorded time between "lights out" and "lights on" should be scored and used for computation of the arousal index.

[1] The definition of sleep stages is not consistent throughout literature. The American Academy of Sleep Medicine (AASM) uses the following terminology: "N1" for sleep stage 1, "N2" for sleep stage 2, "N3" for sleep stage 3 and 4, "R" for REM-sleep and "W" for periods of awakening.

Note 4. The 10 seconds of stable sleep required prior to scoring an arousal may begin in the preceding epoch, including a preceding epoch that is scored as stage W.

Note 5. An arousal may still be scored if it immediately precedes a transition to stage W. That is, both arousal and transition to wake are scored." (Berry, et al., 2016)

Even though this definition is widely used by sleep experts for visual scoring it turns out not to be very precise when it comes to automatic analyzation of EEG signal and leaves space for interpretation.

The scoring of arousals during the rapid eye movement (REM) phase requires a concurrent increase in the submental electromyography (EMG) lasting for at least 1s. The considered sleep laboratory uses the chin EMG for submental EMG. Spontaneous arousals in NREM sleep may occur without an increase in the submental EMG amplitude. This extra condition for REM sleep was added because spontaneous bursts of alpha rhythm are a quite common occurrence in REM (but not NREM) sleep.

Sometimes it's difficult to distinguish between an arousal and the transition to wake. In the work *An Introduction to Sleep and Polysomnography* (Carney MD, Geyer MD, & Sachin PHD, 2012), the authors suggest that an arousal is only scored when the shift in EEG frequency is followed by at least 10 continuous seconds of any stage of sleep. On contrary the AASM says in a note related to the scoring rule of arousals that an arousal may still be scored if it immediately precedes a transition to wake. Due to different opinions on this topic and the lack of a clear difference between arousals and the transition to wake, the 30s preceding a wakefulness will be excluded from arousal detection.

The arousal index is a common magnitude to describe the frequency of arousals and is usually computed as the number of arousals per hour of sleep.

Association of an arousal and a PLM event

An arousal and a limb movement that occur in a PLM series should be considered associated with each other if they occur simultaneously, overlap, or when there is <0.5 seconds between the end of one event and the onset of the other event regardless of which is first. (Berry, et al., 2016)

Figure 2 visualizes the association of an arousal and a PLM.

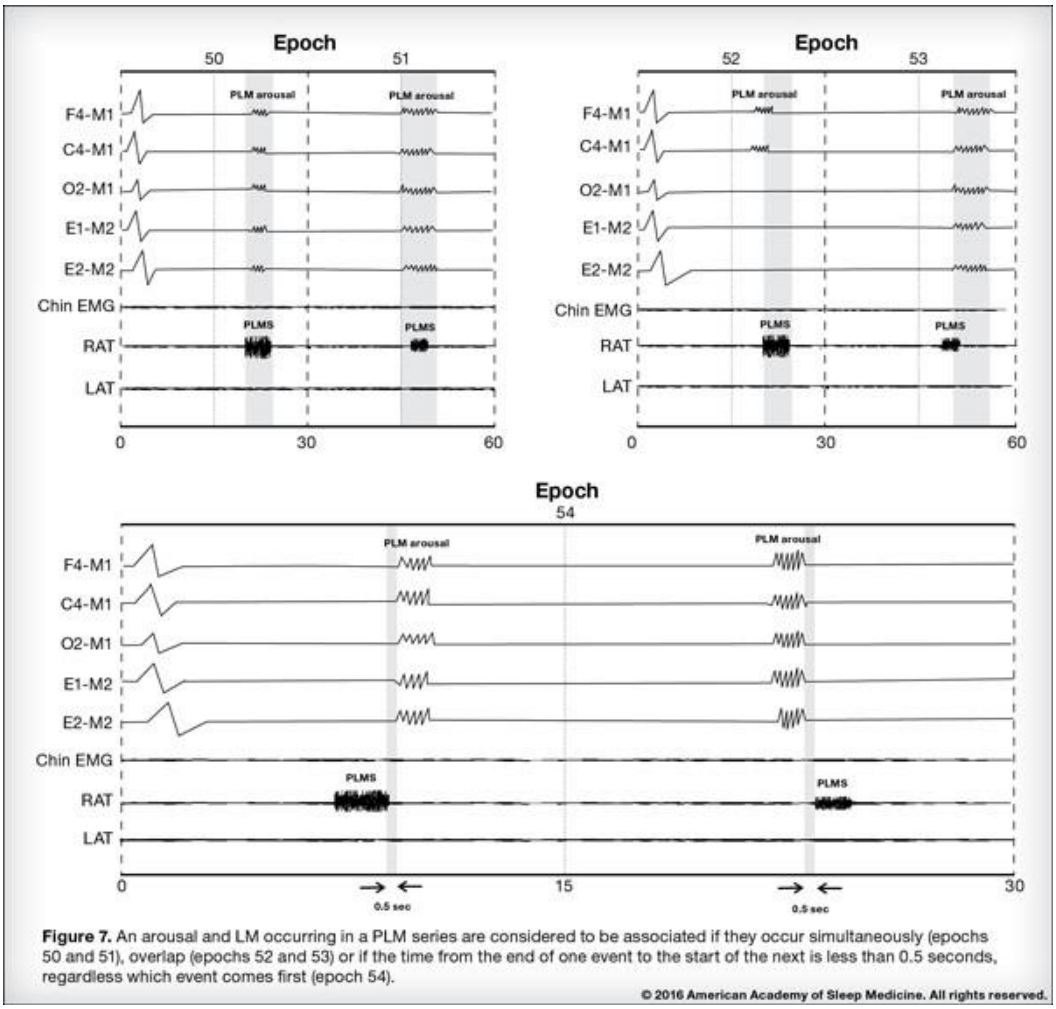


Figure 7. An arousal and LM occurring in a PLM series are considered to be associated if they occur simultaneously (epochs 50 and 51), overlap (epochs 52 and 53) or if the time from the end of one event to the start of the next is less than 0.5 seconds, regardless which event comes first (epoch 54).

Figure 2: Association of a PLM event and an arousal (Berry, et al., 2016)

1.4 State of the art

To score arousals polysomnographic recordings must be analyzed, especially the EEG signal. Traditionally some special ink-writing pens were used to produce 30s epochs on paper, which later were analyzed manually by an expert. Therefore, sleep was staged in epochs of 30s. Later this convention of scoring sleep stages and other events in 30s epochs was also used in digital analysis and is still standard. Nowadays the scoring of arousals is done partly automatically and partly manually by experts, which is time-consuming and quite subjective. For that reason, a reproducible and accurate automation of this process is highly desirable. As later described, the used data is provided by the *AKH Wien*, where the scoring of arousals is still done manually.

During a meeting with sleep medicine experts Mrs. Böck and Prof. Seidl at AKH Wien Sleep Laboratories it was possible to gain an insight into the daily work of the analysis of polysomnographic data. In the following a short summary of important points that should be mentioned for better understanding is provided:

- The annotation of the data used in this work is done by following the definition of the AASM and by using two central EEG derivations C3A2 and C4A1 and the chin EMG for the scoring of arousals during REM sleep. An arousal doesn't need to be clearly seen in both channels.
- Additional channels are used for the scoring of "critical arousals":
    - ECG: An arousal seems to come along with an accelaration in ECG singal.
    - Respiratory Signals: An arousal seems to come along with irregularities in breathing
    - EMG: An arousal seems to come along with some leg movements
- If a K-Komplex or a Delta-Artifact with similar shape directly precedes an arousal it is not counted as part of the arousal.
- Due to medical reasons arousals scored during wakeness have less importance for the microanalysis of sleep and the further diagnosis of sleep disorders. Therefore most of the time arousals during waking phase are not scored.
- Note 5 of the definition of an arousal according to the AASM says that the transitions to wake can also be scored as an arousal. This has to be seen critically because the same event would count to the arousal index as well as to the number of transitions to waking phase during the night. Hence, most of the time transitions to wake are not additionally scored as arousals.
- The end of an arousal is not defined by the AASM and is therefore rather subjective. Sleep experts say that an arousal ends when the EEG signal returns to "normal sleep waves".
- The maximum duration of an arousal depends on its start within a 30s epoch.
    - Within one 30s epoch an arousal can last a maximum of 15s. If it lasts longer the whole epoch is scored as wake.
    - Within two 30s epochs an arousal can last a maximum of 30s starting from the middle of the first epoch and ending in the middle of the second epoch. If is lasts longer one or both epochs are scored as wake.
- Arousal scoring during REM sleep requires a significant increase in chin EMG. The increase shall be seen at the start of an arousal.
- The exact definition of the 10s stable sleep preceding an arousal is a discussion point. But it's clear that there can't be another arousal within the first 10s after the end of an arousal.

1.5 State of knowledge

The aim of this chapter is to provide an overview of already existing publications on this topic. A part of the studies mentioned in this section specialize on a very specific group of patients, for example patients with obstructive sleep apnea syndrome or with periodic limb movement syndrome. The publications are analyzed with special focus on preprocessing of the signal, regarding segmentation and filtering, the extracted features and the used algorithms. Finally, the results of the mentioned studies are provided for comparison reasons.

Automatic Detection of Sleep Arousal Events from Polysomnographic Biosignals
(Shahrbabaki, Dissanayaka, Patti, & Cvetkovic, 2015)

In the work of (Shahrbabaki, Dissanayaka, Patti, & Cvetkovic, 2015) 9 subjects were analyzed and out of these, 4 suffered from Obstructive Sleep Apnea, 1 from PLMD and the remaining 4 subjects were healthy. The aim of this paper is to develop an algorithm for automatic detection of sleep arousals without distinguishing different arousal types or sleep disorder groups. Nevertheless, different feature

sets are extracted for the detection of special arousal types, as for instance arousals related to respiratory events, are extracted. After feature extraction, a K-nearest-neighbors (KNN) algorithm is applied to classify epochs of 30s in arousal epochs and no-arousal epochs.

This study makes use of more than just the EEG-signal. In addition, the left and right electrooculogram, submental electromyogram, electrocardiography, leg movement actigraphy, nasal and oral airflow, snoring sound, saturation of oxygen and chest and abdomen breathing effort are used.

The signal is segmented in 30s epochs, a common convention in sleep medicine, and bandpass filtered through a Butterworth filter with second order cut-off frequency from 0.4 to 40Hz. To obtain the power spectral values the epochs are transformed to frequency domain with Welch's algorithm (a sequential Fast Fourier Transformation). The power spectral values are categorized into the following EEG bands: $\delta\ (0.5-4\ Hz), \theta\ (4-8\ Hz), \alpha\ (8-15\ Hz)\ and\ \beta\ (16-32\ Hz)$.

Features:

- *EEG features:*
  - For each epoch of 30s and each of the three channels used, the percentage of signal power generated by a special frequency band was calculated.
  - Within each epoch the ratio of α- and β- power (signal filtered from 8-40 Hz) of a 10s window and a consecutive 3s window is calculated for each moment. As features the minimum and maximum as well as the area under the resulting curve of the ratio in signal power is calculated.
- *EMG features:*
  - For each epoch of 30s the root mean square of the submental EMG signal is calculated.
  - In addition the EMG was transformed into frequency domain using Welch's algorithms and the center of power spectral values was calculated.
- *Leg movement features:* The root mean square of the leg movement time series was calculated and used as a special feature for the detection of movement related arousals
- *Respiratory features:* The airflow time series was transformed into frequency domain and three features such as arithmetic mean, standard deviation and maximum of power spectral values were calculated.
- *Heart rate features:* Peaks of ECG signal were detected using an algorithm proposed by Murthy et al. and statistical values for heart rate were stored.

Detection algorithm:

A feature matrix with n rows for the number of 30s epochs and 32 columns for the extracted features was built and classification was done with a K-Nearest-Neighbor algorithm. This algorithm classifies 30s epochs in arousal epochs or no-arousal epochs. An exact localization of the arousal within this epoch is not performed.

Validation:

Common statistical performance indicators like Sensitivity, Specificity and Accuracy were used for Leave-One-Out-Cross validation. On average arousal epochs of 30s could be scored with an average of 79% in sensitivity, 96% in specificity and 94% in accuracy.

A method for the Automatic Detection of Arousals during sleep

(De Carli, Nobili, Gelcich, & Ferrilo, 1999)

The aim of this work is to develop an algorithm to mark segments as arousal segments using two EEG and one submental EMG derivation.

In the study of (De Carli, Nobili, Gelcich, & Ferrilo, 1999) the EEG-data is transformed with a wavelet transformation (uses a family of functions to decompose the signal) of thirty-two second overlapping epochs. The signal power was evaluated with a time resolution of 0.125s, for six frequency bands: $slow\ \delta\ (0-0.5Hz),\ \delta\ (0.5-4Hz),\ \theta\ (4-8Hz),\ \alpha\ (8-12Hz),\ \sigma\ (12-16Hz)\ and\ \beta\ (16-64Hz)$.

Features:

To detect shifts in EEG frequency, especially an increase in theta, alpha and/or beta activity, the following indices are used for every 0.125s basic epoch:

- *EEG features:*
  - Signal power within the 6 frequency bands in epochs of 0.125s
  - A long term and a short term moving average
    - First six indices are the ratio between short term and long term average in each band (indicating variation within each band).

      Special features in addition:

    - Ratio between short term and long term mean (central) frequency
    - Ratio between δ and α plus β power for short-term and long-term average (highlights presence of slow wave sleep)
    - Ratio between short-term and long-term α relative power (highlights variations in alpha activity)
    - Ratio between long-term α plus *slow-*δ and θ plus δ power (could indicate awakening)
    - Ratio between σ und α plus β power (highlights sleep spindles)
    - Ratio between β and δ variations
- *EMG features:*
  - A long term and a short term moving average of the submental EMG signal were computed and compared as in the EEG analysis.

Detection algorithm:

After marking and weighting of overlapping EEG and EMG events a linear discriminant function was estimated by maximizing sensitivity and specificity on the training set. Possible arousals were marked when the discriminant function remained positive for 3-30s. The scoring of arousals was done with respect to the mean value of the discriminant function.

Validation:

The discriminant function is tested and estimated on a test set of patients with different disorders. Afterwards the sensitivity and the selectivity are calculated and the computer program as well as a second expert are compared to the original expert annotations. Considering only definite arousals (for

more detail consider the publication of (De Carli, Nobili, Gelcich, & Ferrilo, 1999)) the sensitivity of the computer program resulted in 88.1% while selectivity was 74.4%.

Automatic Detection of Micro-Arousals (Agarwal, 2005)

The method presented in this paper is based on three basic steps: the segmentation in variable length epochs (adaptive segmentation), the spectral feature extraction and the identification of EEG epochs with potential arousals using statistical methods and decision rules.

The central lead EEG is filtered in the extended β-band (16-40 Hz) and the occipital data is filtered in the α-band (7.5-12 Hz). Using these filtered data, the signal is segmented (in segments with a minimum length of 3s) according to a non-linear energy operator, which enhances bursts of α- and β-frequencies. Features are calculated on segments of different length. The idea is to calculate a new frequency weighted energy operator and apply it consecutively to two parts of a rolling window and calculate the difference. High differences in frequency between the two windows are pointed out at the maxima of the new curve.

Features:

- Total power of α- and extended β- band in each segment
- Find the maximum absolute amplitude of each considered band in each segment

Detection algorithm:

In order to select the correct segments containing candidates for arousals (CMA) the following steps are done:

- Remove segments containing large amplitude artifacts and all segments not occurring in valid sleep stage (1-4 and REM)
- Calculate the mean α/β power and stay with all the segments where the mean of α/β power is K times the standard deviation greater than the entire mean.

The final selection for arousal segments is done as follows:

- Remove all CMA which are shorter than 3s and occur at least 10s after the last arousal (mandatory 10s of stable sleep) and check if the whole CMA appears in a valid sleep stage
- Stay with the CMA in which either α or β power that is at least 2 times greater than the left or right neighboring segment

Validation:

The method was compared to three expert annotations. Compared to two of them the algorithm showed more consistent results with sensitivity from 70.1% to 82.2% and specificity from 56.6% to 72.4%.

Identification of Electroencephalographic Arousals in Multichannel Sleep Recordings
(Álvarez-Estévez & Moret-Bonillo, Identification of Electroencephalographic Arousals in Multichannel Sleep Recordings, 2011)

The methods presented in this paper are based on two central EEG channels (C3A2, C4A1) and a submental EMG signal. After extracting several features four different classifiers, namely Fisher's linear and quadratic discriminant, a support vector machine (SVM) and an artificial neural network (ANN), are compared. By using machine learning models, the method is no longer dependent of a set of prefixed thresholds.

It must be mentioned that there is a second publication of the same authors available applying mostly the same methods on Sleep Apnea patients. (Álvarez-Estévez & Moret-Bonillo, Model Comparison for the Detection of EEG Arousals in Sleep Apnea Patients, 2009)

In order to prepare the signal for feature extraction it is bandpass filtered within characteristic bands: $\delta\ (0.5-3\ Hz), \theta\ (4-7\ Hz), \alpha\ (8-12\ Hz), \beta\ (13-30\ Hz), \sigma\ (12-14\ Hz)$, as well as a joined $\alpha\beta - band\ (8-30\ Hz)$.

Features:

The main idea of feature extraction is to get an information of the evolution of the signal by comparing the signal power of a 10s window, representing the prior information, with the power of an immediately consecutive 3s window for each second of the signal.

- *EEG-features:*
  A new curve $e_{\alpha\beta}$ is calculated (1Hz resolution) by comparing the $\alpha\beta$ signal power (power of the signal filtered within the band $8-30Hz$) within the two consecutive windows, and a possible event is marked when $e_{\alpha\beta}$ is greater than zero:
  - Area under $e_{\alpha\beta}$ and duration of the possible event
  - Maximum value of $e_{\alpha\beta}$ during the possible event
  - Area under curve, maximum and minimum for each power change curve e obtained for each single frequency band
- *EMG-features:*
  A new curve a is calculated (1Hz resolution) by comparing the amplitude within the two consecutive windows (here 30s and 3s), and a possible event is marked when a is greater than zero:
  - Area over a certain threshold and duration of the possible event
  - Maximum value of a during the possible event
- *Contextual features:*
  Binary features marking if a possible event is marked in one of the EEG-channels or in the submental EMG.

Detection algorithm:

For classification, the signal is segmented in classifiable epochs of 30s and described by feature vectors of the possible events within the epoch. Four different classifiers are tested and their ability to classify epochs of 30s as arousal or no-arousal epochs is compared:

- Fisher's linear discrimimant
- Fisher's quadratic discriminant
- Support vector machine with radial basis kernel
- Artificial neural network

Validation:

The validation was carried out on balanced and unbalanced test and training sets (for more detail see (Álvarez-Estévez & Moret-Bonillo, Identification of Electroencephalographic Arousals in Multichannel Sleep Recordings, 2011)) and resulted in a mean sensitivity of 86% and specificity of 76% within the finally chosen training and test sets.

Automatic artifacts and arousals detection in whole-night sleep EEG recordings
(Wallant, et al., 2016)

The arousal detection in the study of (Wallant, et al., 2016) only plays a minor role next to the artifact detection. Signal is filtered with a Butterworth filter of order 3 and low- and high-pass filtered with a forward-backward filter. The segmentation is done fixed in 20-30s windows and within these in 1s-epochs. Power spectral density was computed from the average of central electrodes, where so called "bad channels" were excluded and the averaged power of frequency bands ($\theta\ (3-7\ Hz)$, $\alpha\ (7-13\ Hz)$, $\sigma\ (11-16\ Hz)\ and\ \beta\ (16-32\ Hz)$) was calculated for each 1s-epoch and stored.

Features:

- *EEG-features:*
  The median of each averaged power vector (1Hz resolution) is calculated for the whole recording and for each scoring window and considered as a fixed threshold to detect abnormal acitivity.
- *EMG-features:* Similar procedure, combination of median, standard deviation and peaks of the signal are extracted as thresholds.

Detection algorithm:

All 1s-epochs in one scoring window are considered as a shift in EEG if their signal power within one of the frequency bands is larger than two times the median within the scoring window as well as greater than the overall median. If three consecutive seconds fulfill this condition an arousal is detected.

Validation:

Since the purpose of this paper is to remove epochs with arousals there is no validation provided.

In addition to the above mentioned and in more detail described studies, there are a few more papers specializing on arousal detection for Sleep Apnea Syndrome as well as arousal detection in Parkinson disease. The methods provided are very similar to the ones in the already described works. For Sleep Apnea patients the following works are referenced: *Automatic EEG Arousals Detection for Obstructive Sleep Apnea Syndrom*e (Sugi, Kawana, & Nakamura, 2008) or (Sugi, Kawana, & Nakamura, 2009),

*Detection of Arousals in Patients with Respiratory Sleep Disorders Using a Single Channel EE*G (Cho, Lee, Park, & Lee, 2005)

A study for Parkinson's disease patients is covered for example in the work: *Detection of arousals in Parkinson's disease patients* (Sorensen, Kempfner, Jennum, & Sorensen, 2011)

1.6 Research question

The practical research was done during a research activity at the AIT Austrian Institute of Technology in corporation with the AKH Wien by analyzing exclusively PSG-data of their sleep laboratories. The analysis will be done by using two sets of annotations: On the one hand the original annotations of the AKH Wien and on the other hand a modified set of annotations with EEG-arousals which show a clear frequency shift and are technically AASM conform without taking other bio-signals into account.

The first aim of this thesis is to develop an automatic detection algorithm to detect arousals in EEG signals following the definition of the AASM American Academy of Sleep Medicine. Several features and statistical measures will be derived to describe the event of an arousal. The algorithm will be tested on the technically modified annotations of the AKH Wien.

The second and clinically motivated aim is to examine whether the EEG-arousals show any relation with annotated leg movements during sleep. To answer this question the original and the technically modified data will be tested.

2 Materials and methods

2.1 Data set

In this thesis, PSG recordings from the AKH-Wien sleep laboratories will be used. More specifically, seven patients with 5 to 9 hours of sleep data recordings that are either diagnosed without sleep disorder or with PMLS are considered.

Identification number of used patients:

- TS100714, TS090714, TC070814, MB200814: diagnosed with Periodic Limb Movements of Sleep (PLMS)
- CP101214, MC250614, MP020714: no sleep disorder diagnosis

Patients with respiratory disorders have been removed because they are not relevant for the clinical question of the relation between leg movements and arousals in this work. Nevertheless, some few remaining respiratory events can occur in the PSG data of all patients. The available EEG data, recorded with a sample rate of 256 Hz, consists in two central derivations: C3A2 and C4A1. Both channels will be used for arousal detection, as well as the chin EMG channel for detection of REM-sleep arousals and the hypnogram to exclude periods of awakening. Figure 3 shows the used data for the arousal detection algorithm. The marked area describes an annotated arousal of 3.079s during sleep stage 2.

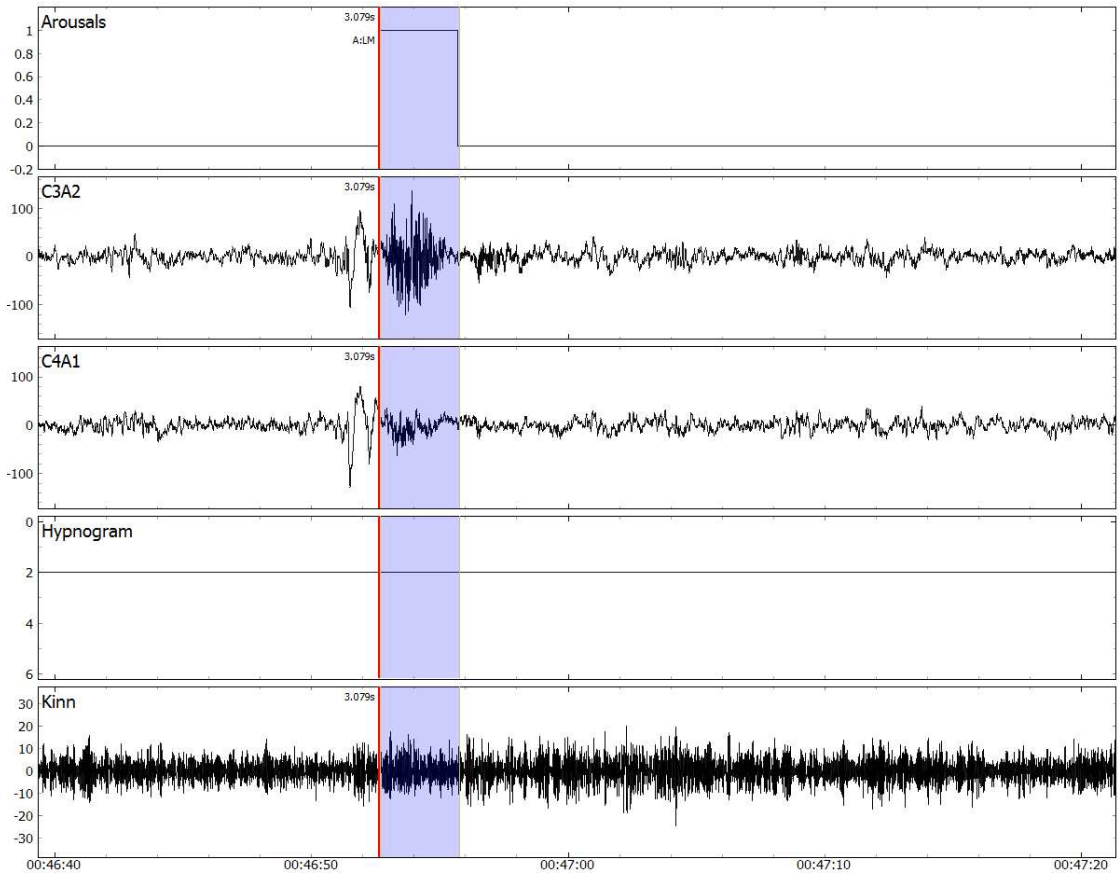


Figure 3: Illustration of the used PSG data for arousal detection, visualized with the "Sleep-Data-Viewer-Tool" of the AIT; 1. Line: Arousal annotations; 2. Line: EEG channel C3A2; 3. Line: EEG channel C4A1; 4. Line: Hypnogram; 5. Line: chin EMG

As described in chapter *State of the art* annotations are mainly done manually by sleep experts. These annotations are clinically approved by considering much more than just the EEG-signal. The experts evaluate the patient by looking at other bio signals like respiratory anomalies, ECG irregularities, leg movements, etc. sometimes even video analysis or the knowledge of the clinical history of the person is considered.

The detection algorithm derived in this work is exclusively based on the EEG signal, the chin EMG and the hypnogram. The features are mainly derived from the definition of the AASM. To evaluate the performance of the algorithm correctly, the annotations needed to be modified to technically clear annotations following the AASM definition without access to other information except for the mentioned data. Although most of the annotations in both sets coincide, there are slight differences, wherefore in the following it will be referred to:

- *Original annotations*
- *Modified annotations*

2.1.1 Original annotations

This term will be used for the original arousal annotations, done manually by a team of sleep experts of the AKH Wien. These annotations may also consider additional information from other bio signals.

2.1.2 Modified annotations

This term will be used for the slightly modified (in case of unclear cases) original annotations by using only the EEG-signal and the definition of the AASM for the annotation decisions. The whole signal of all 7 recordings was checked with the *"Sleep-Data-Viewer-Tool"* (see Figure 3) of the AIT and newly annotated. Since the definition of an arousal is rather vague and the annotations of AKH Wien are made professionally manually by experts, the intention was not to change the annotation set too much. Therefore, arousals were only added in clear abrupt frequency shift cases that don't let any doubts and likewise arousals were only removed from the annotation set if the criteria of the definition of the AASM was not clearly fulfilled.

For the documentation of the modified annotation set, each arousal was classified in either "unclear" or "clear" arousal and one of 10 reasons was added to the comments while annotation process:

- **"Expert annotation":** Describes an annotation done by a sleep expert and used without changes for the modified set.
- **"Modified expert annotation":** Describes an annotation done by a sleep expert but changed in start or end time.
- **"Added annotation":** Describes an annotation that was added to the modified set but wasn't in the original set.
- **"Patient awake":** Describes an annotation that occurs partly or completely when hypnogram was set zero. These got removed for the modified annotation set because epochs of awakening are not considered in the algorithm, as described later.
- **"No frequency shift":** Describes an annotation that doesn't show a clear frequency shift and was probably annotated because of other criteria, which are not available in the following detection algorithm.

- **"Confusion with artifact":** Describes an annotation that shows a very high amplitude, mostly for a very short time, that is more likely to be an artifact than an arousal.
- **"Confusion with delta-artifact or K-Komplex":** Describes an annotation that shows high delta frequencies or only consists in a sharp wave similar to a K-Komplex. Since a shift in delta frequencies is not considered to be an arousal these are also removed.
- **"Confusion with sleep spindle":** Describes an annotation that shows high sigma frequencies that are a criteria for a sleep spindle and may have been confused with alpha frequencies. Since sleep spindles need to be distinguished from arousals these are also removed.
- **"Freq shift shorter than 3s":** Describes an annotation that shows a frequency shift but it lasts shorter than 3s.
- **"No stable sleep 10s before":** Describes an annotation that is not fullfilling the 10s of stable sleep before the arousal because of an other arousal or because of frequency shifts shorter than 3s that are not considered as an arousal.

The modified set consists of arousals classified as "clear". These are all with one of the following comments:

- **"Expert annotation"**
- **"Modified expert annotation"**
- **"Added annotation"**

Table 1 and Table 2 give an overview of the difference between the original and the modified annotation set.

	TS100714	TS090714	TC070814	MB200814
# arousals in original set	*238*	*235*	*113*	*115*
# arousals in modified set	*205*	*201*	*79*	*83*
% of arousals in the original set that got removed	*17%*	*18%*	*30%*	*30%*
% of arousals in the modified set that are added annotations	*4%*	*5%*	*0%*	*2%*

Table 1: Statistics of original and modified annotation set of PLMS patients

	CP101214	MC250614	MP020714
# arousals in original set	*15*	*78*	*76*
# arousals in modified set	*11*	*68*	*53*
% of arousals in the original set that got removed	*27%*	*17%*	*32%*
% of arousals in the modified set that are added annotations	*0%*	*4%*	*2%*

Table 2: Statistics of original and modified annotation set of patients without sleep disorder diagnosis

2.2 Methods for arousal detection

The aim of the following chapter is to explain the functionality of the arousal detection algorithm and to derive the methods used in each step. As already mentioned in the section *Data set*, the automatic detection will be developed and evaluated with a modified data set of annotations. Arousals that can't be clearly annotated without the information of other channels got removed. The appearance of an arousal may take various forms, depending on the patient and many other factors like leg movements, respiratory events, sleep stage, etc.

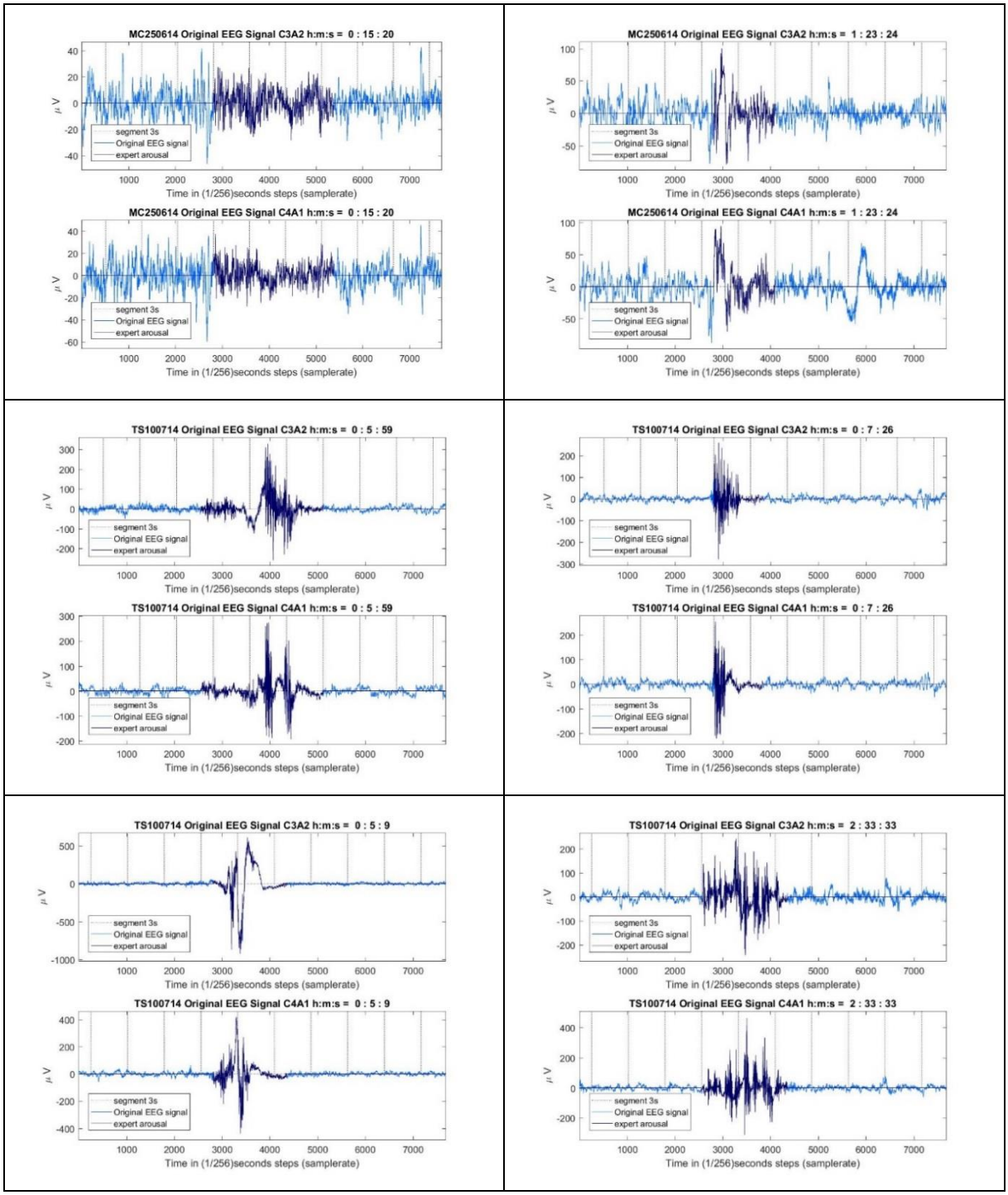


Figure 4: Various forms of arousal appearance in two channels

Figure 4 illustrates that some arousals may be easier to identify than others. Most of the time there is also an amplitude shift noted next to the frequency shift.

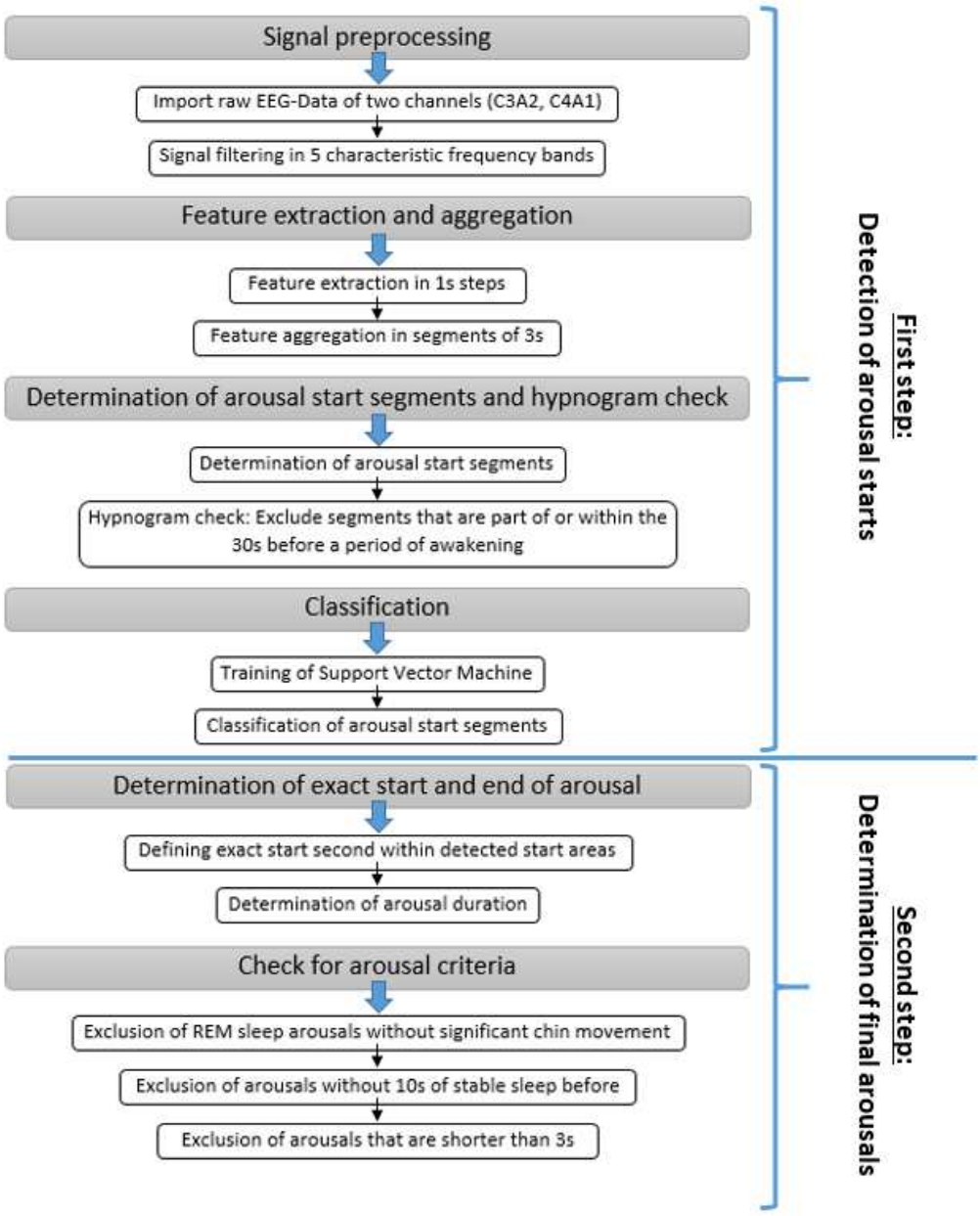


Figure 5: Overview of arousal detection algorithm

Figure 5 gives an overview about the arousal detection process. In the following sections the details of the detection process are explained and methods are derived.

2.2.1 Signal preprocessing

2.2.1.1 Signal filtering

As described in *Polysomnographic terms and definitions* in sleep medicine five characteristic frequency bands are analyzed. Hence, it is necessary to bandpass filter the EEG signal. The implemented filter is a window-based finite impulse response (FIR) filter constructed with a Hamming-window. The following methods are derived based on the book *Zeitdiskrete Signalverarbeitung* (Oppenheim, Schafer, & Buck, 2004) and the script *Stationäre Prozesse und Zeitreihenanalyse* (Scherrer, 2015).

Construction of a FIR-bandpass filter

Let $(x_t|t \in \mathbb{Z})$ be the input signal and $(y_t|t \in \mathbb{Z})$ the filtered signal. A linear, dynamic filter is a function of the form:

$$(x_t) \mapsto A(x_t) \coloneqq \left(y_t = \sum_{j=-\infty}^{\infty} a_j x_{t-j} \right), with\ (a_j \in \mathbb{R}^{mxn} | j \in \mathbb{Z})$$

A filter with $(a_j|j \in \mathbb{Z})$ is commonly called:

- Dynamic, if $a_j \neq 0$ for at least one $j \neq 0$
- Static, if $a_j = 0$ for all $j \neq 0$
- Causal, if $a_j = 0$ for all $j < 0$
- Anti-causal, if $a_j = 0$ for all $j > 0$

The output y_t doesn't have to exist for a filter with an infinite number of filter coefficients. To construct a filter with a well-defined output, one has to impose certain assumptions on coefficients and input. If the input (x_t) is bounded, $|x_t| < c$, and the filter coefficients are absolutely summable ($\rightarrow l1 - filter$), then the output y_t is well-defined. One option to fulfill the condition, that filter coefficients are absolutely summable, is to create a filter with a finite number of non-zero filter coefficients. This is the main idea of constructing FIR-filters with the window-method. A non-causal FIR-filter of order q with 2q+1 filter coefficients (finite number) is defined as follows:

$$(x_t) \mapsto A(x_t) \coloneqq \left(y_t = \sum_{j=-q}^{q} a_j x_{t-j} \right), \quad with\ \ t > q+1\ and\ t < T-q$$

When analyzing filter in frequency domain the transferfunction of the filter A with coefficients $(a_j|j \in \mathbb{Z})$ can be defined as the discrete Fourier transformation of the filter coefficients:

$$\lambda \in [-\pi, \pi] \mapsto a\left(e^{-i\lambda}\right) = \sum_j a_j e^{-i\lambda}$$

The function $\lambda \mapsto |a(e^{-i\lambda})|$ is called the gain of the filter. The gain of the filter can tell which frequencies are enhanced and which are damped:

- If $\left|a\left(e^{-i\lambda}\right)\right| < 1$ frequency λ is damped
- If $\left|a\left(e^{-i\lambda}\right)\right| > 1$ frequency λ is enhanced

For the further analysis of the characteristic frequency bands of the EEG-signal, it is necessary to construct a bandpass filter, to "let through" only a specific frequency range of the signal. The green area in Figure 10 and Figure 11 illustrates a perfect bandpass filter, with a filtergain that equals one within the desired frequency band and is zero elsewhere. Unfortunately, there is no $l1 - filter$ with such a transferfunction, because this transferfunction is not continuous. The aim is to find a filter with a well-defined output and a transferfunction that is relatively close to the ideal one.

Let $(a_j | j \in \mathbb{Z})$ be the filter coefficients of the ideal bandpass filter. A common method to construct a FIR filter is to construct a window $w(j)$ defined as:

$$w(j) = \begin{cases} w_j, & j \in [-q, q] \\ 0, & else \end{cases}$$

Figure 6 shows two possible choices of window functions.

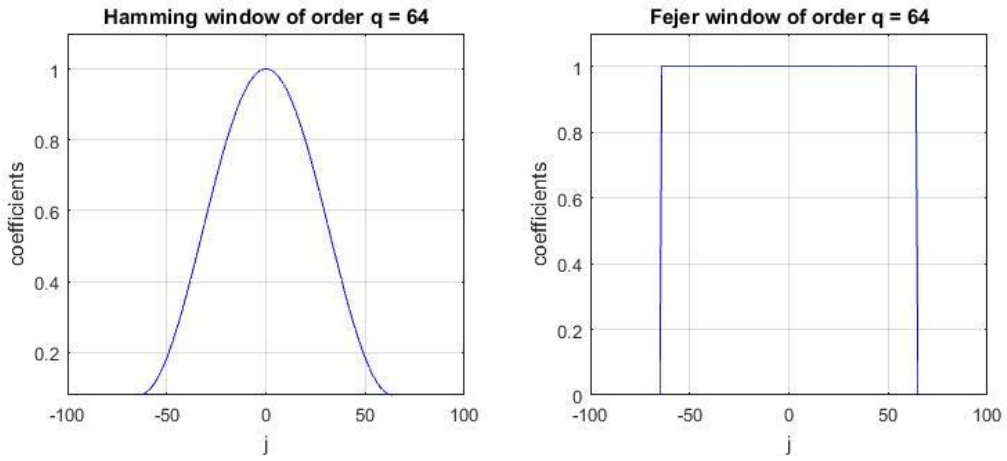


Figure 6: Illustration of Fejer-window (rectangular window) and Hamming window of order q = 64

The filter coefficients (a_j^w) of the FIR filter then are defined as the product of the filter coefficients (a_j) with the window function $w(j)$, $j \in \mathbb{Z}$:

$$a_j^w = a_j \cdot w(j)$$

The transferfunction of the FIR filter with coefficients $(a_j^w = a_j \cdot w(j))$ is the convolution of the transferfunction $a(e^{-i\lambda})$ and the discrete Fourier transformation of the window function $w(e^{-i\lambda}) = \sum_{j=-q}^{q} w(j)\, e^{-i\lambda}$. It can be written as:

$$a^w(e^{-i\lambda}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} a(e^{-i\theta}) w(e^{-i(\lambda-\theta)}) d\theta$$

When cutting the coefficients of a filter to finite length with a window function, one can observe a phenomenon called leakage effect. It describes that when transforming the signal into frequency domain, frequency components, that would not appear if the filter with filter coefficients of infinite length were used, inevitably emerge. The aim is choosing a window function that reduces this effect as much as possible.

As visualized in Figure 6 on the right side, one of the simplest versions of a window function is the Fejer-window (rectangular window) which is defined as:

$$w_f(j) = \begin{cases} 1, & j \in [-q, q] \\ 0, & else \end{cases}$$

Another window function was proposed by the US mathematician R. W. Hamming. In order to minimize the side lobes closest to the main lobe (see Figure 7-11) the hamming window is defined as

$$w_h(j) = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\pi j}{(2q+1)-1}\right), & j \in [-q, q] \\ 0, & else \end{cases}$$

with $(2q+1)$ the width (in samples) of the symmetrical window function.

Figure 7 and Figure 8 show the modulus of the Fourier transformation of both presented window functions in two different resolutions as well as in dB for filter orders $q = 64$ and $q = 256$.

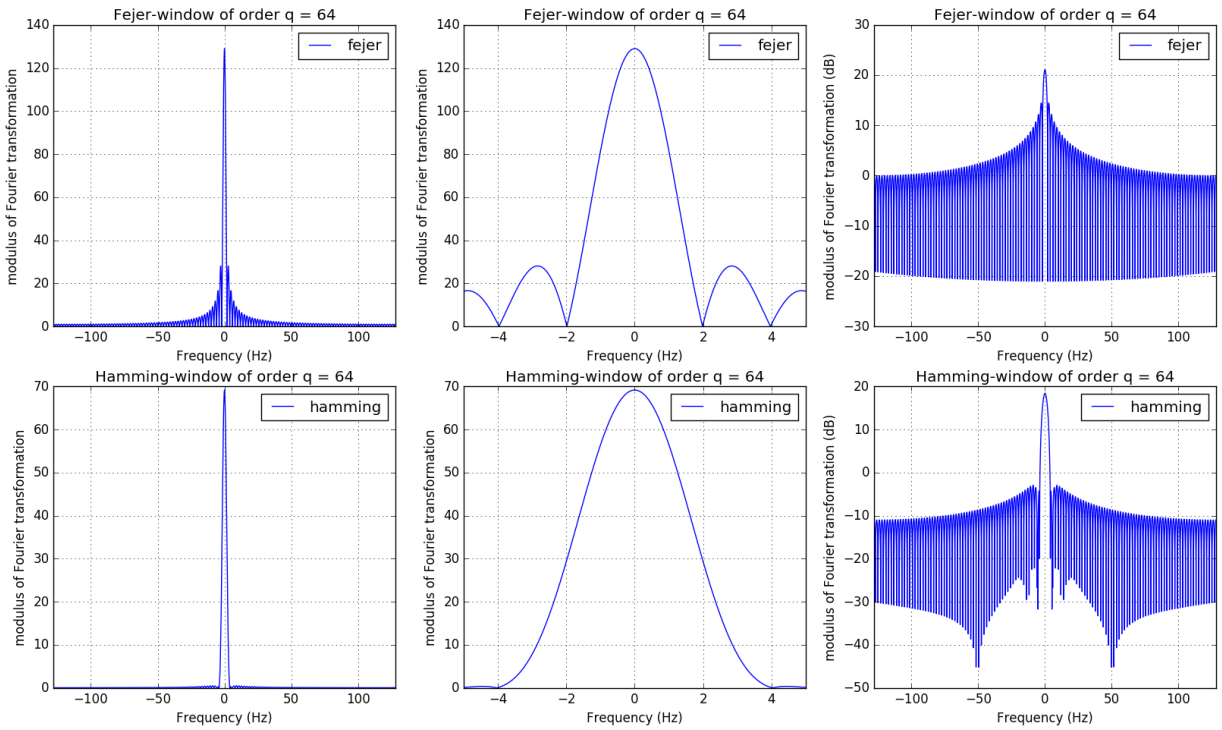


Figure 7: Modulus of the Fourier transformation of both presented window functions of order q=64

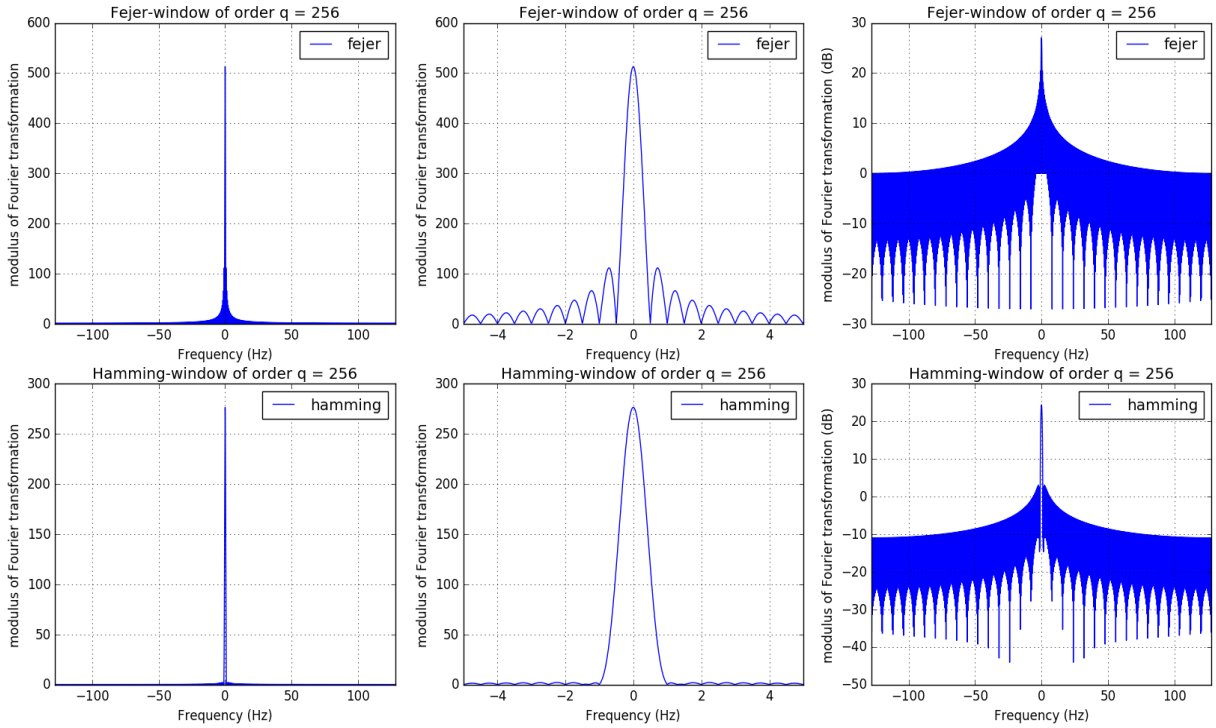


Figure 8: Modulus of the Fourier transformation of both presented window functions of order q=256

Clearly, for the Fejer-window the coefficients within the non-zero interval $[-q, q]$ are the ideal filter coefficients. Figure 9 shows the filter coefficients for filter order 64, 128 and 1024 of the FIR-bandpass filter (4-7Hz) calculated as the product of the coefficients of the ideal filter with the ones of the window function.

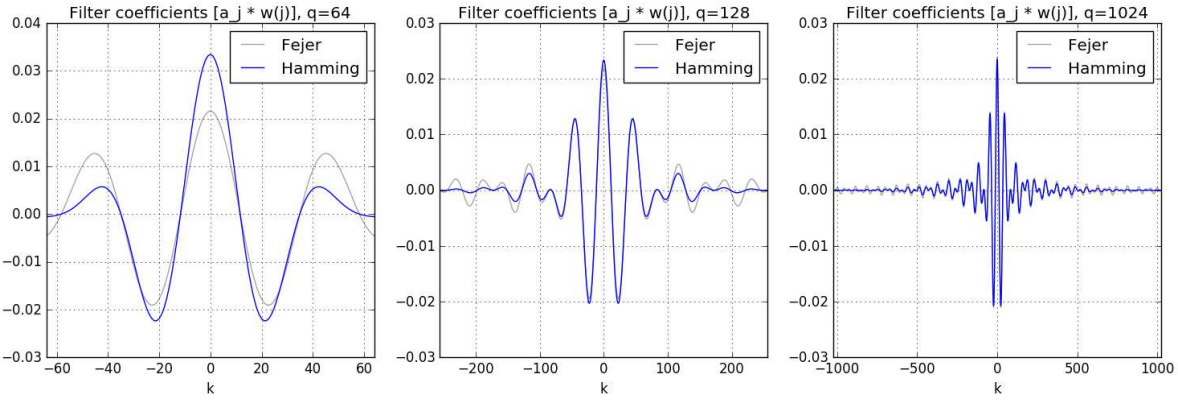


Figure 9: Filter coefficients of the FIR bandpass filter for 4-7 Hz constructed with window functions

When visualizing the gain of the FIR bandpass filter for the frequency range 4-7 Hz with coefficients (a_j^w), the different characteristics of the window functions can be observed. Figure 10 and Figure 11 show the filter gain of the bandpass filters and it can be observed that the main lobe of the Hamming window is wider than the one of the Fejer-window but the side lobes closest to the main lobe of the Hamming-window are smaller.

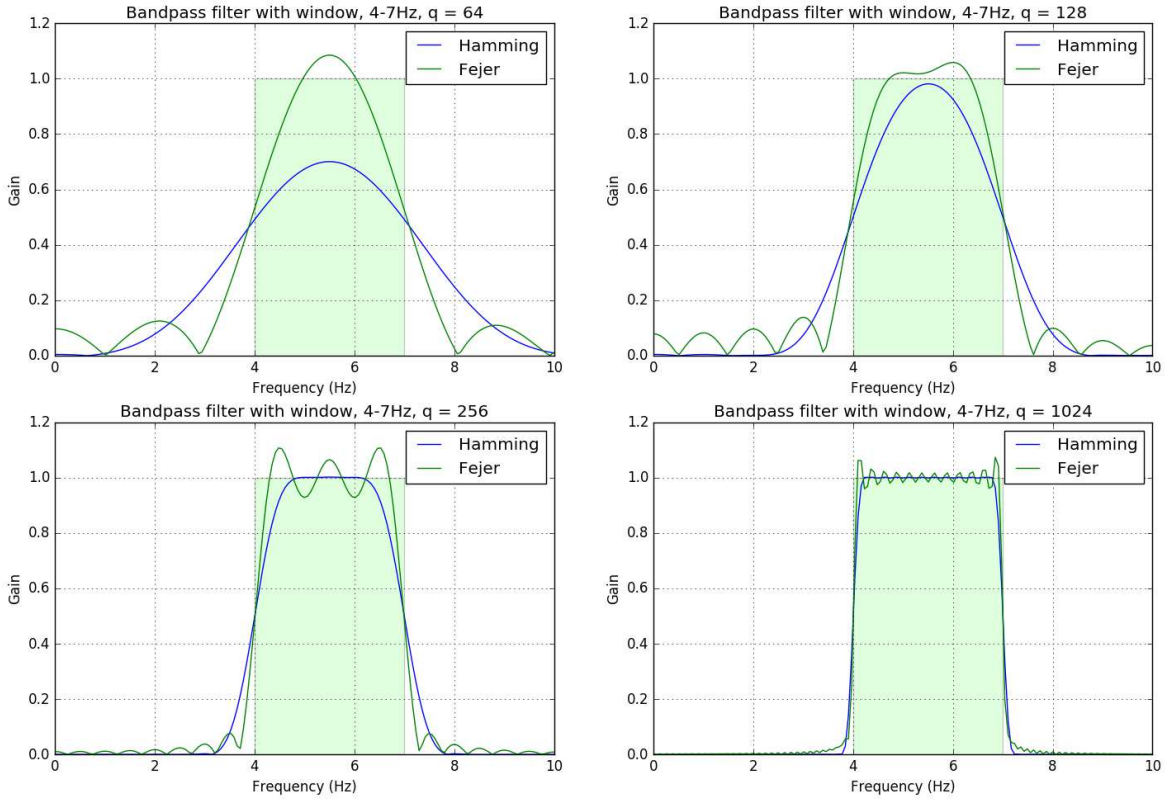


Figure 10: Filter gain of the bandpass filter for 4-7 Hz constructed with two different window functions

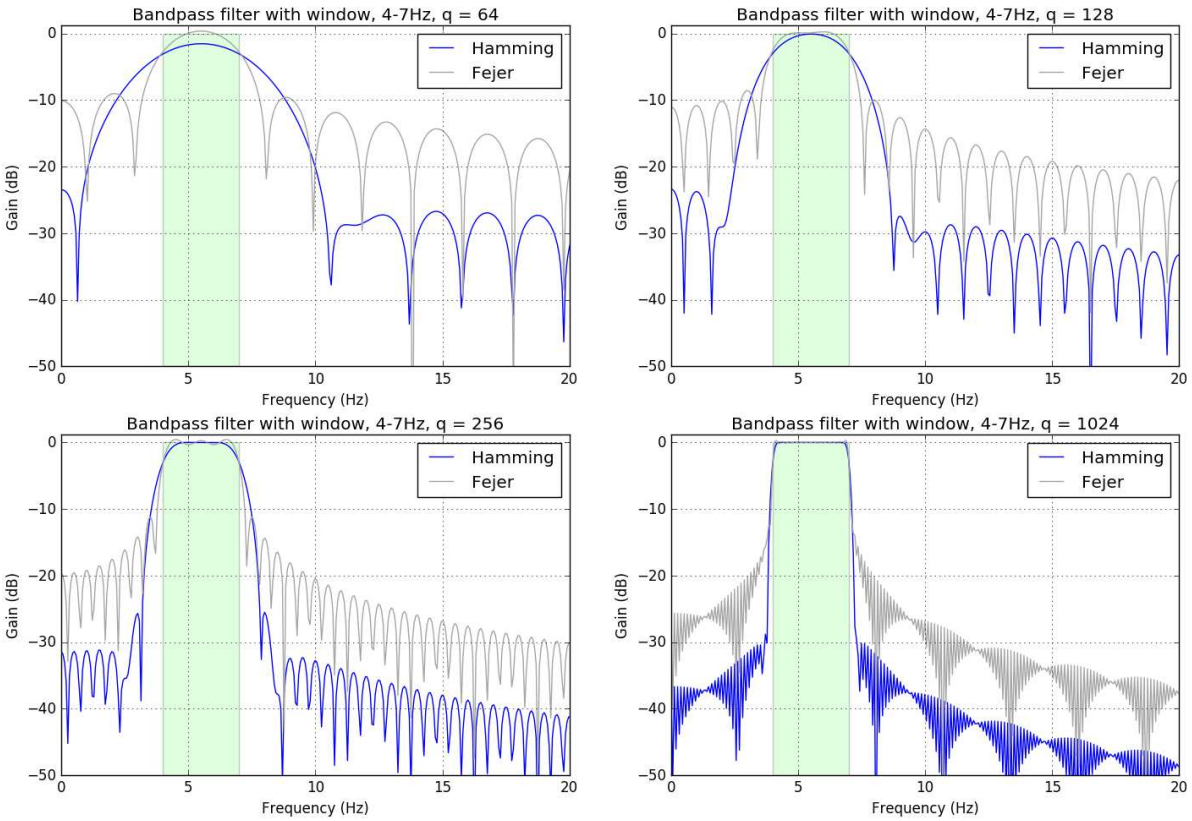


Figure 11: Filter gain of the bandpass filter for 4-7 Hz in dB constructed with two different window functions

The higher the chosen filter order is, the more the precision of the bandpass filter increases. But the drawback is that filtering the signal with a high number of coefficients means that the filtered signal y_t depends on a high number of future and past data points and that could influence time critical detection problems. Therefore, to guarantee enough precision for the detection problem, a filter order of $q = 128$ is used, which corresponds to half a second of time (past and future data points for a sample rate of 256Hz).

Signal filtering in Python

The filter coefficients for the bandpass filter constructed with a hamming window are calculated in **python** with the function `filtwin` out of the signal processing package `scipy.signal`. For the calculation of the filtered signal the function `lfilter` in **python** is used, which implements a causal filter of the form:

$$\tilde{y}_t = \sum_{j=0}^{2q} b_j x_{t-j}\,, \quad t > 2q \qquad (1)$$

In order to compute the output of a non-causal (two-sided) FIR filter of the from

$$y_t = \sum_{j=-q}^{q} a_j x_{t-j}\,, \quad t > q+1 \; and \; t < T-q \qquad (2)$$

by the function `lfilter` we simply set $y_t = \tilde{y}_{t+q}$ and $b_j = a_{j-q}$.

In practice, to compute the output (y_t) of filter *(2)* out of filter *(1)*, realized with **python**, the filtered data needs to be shifted back q steps by cutting the first q points and adding q zeros at the end of the signal. Since each EEG-recording has millions of data points the missing filtered data at the beginning and at the end doesn't influence any results.

Figure 12 shows a 12s interval of EEG-signal of the central derivation C3A2 without any irregularities filtered into the 4 characteristic frequency bands (δ-band, θ-band, α-band and β-band). The sigma waves are not of great interest here, because they would indicate sleep spindles and must be excluded from arousal detection.

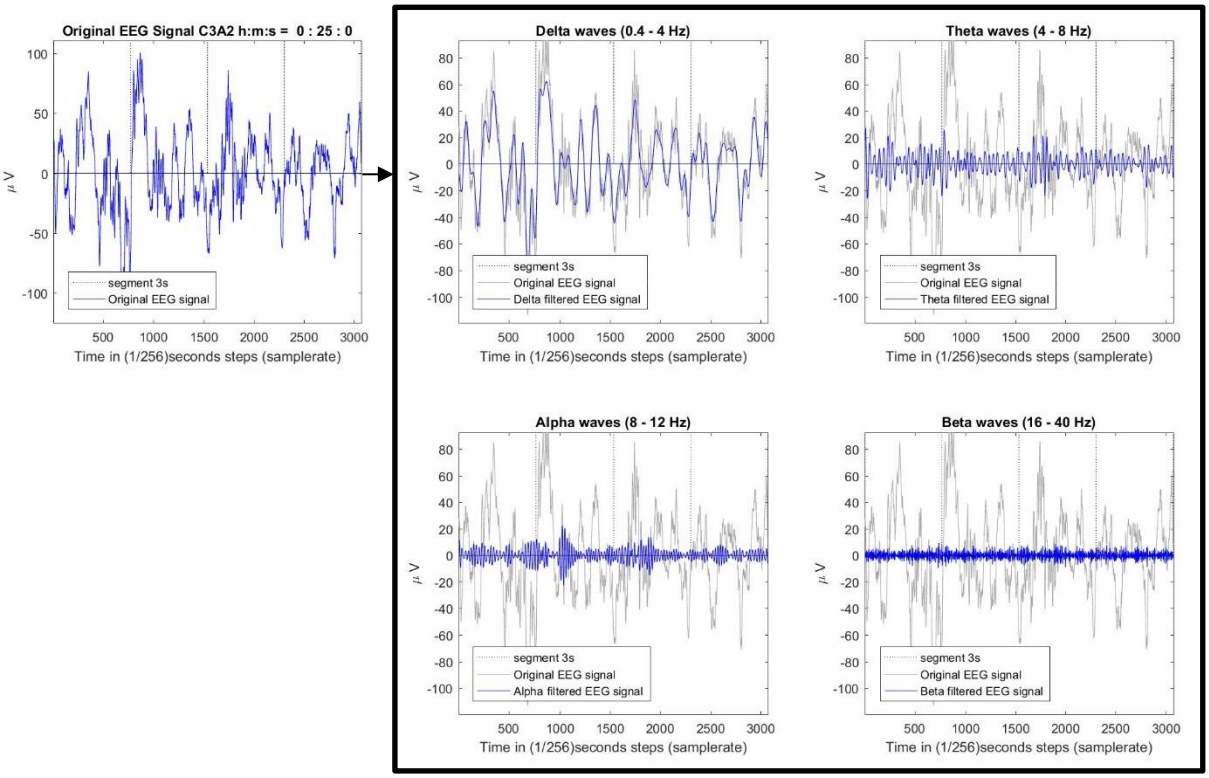


Figure 12: Original signal bandpass filtered with a FIR-Filter constructed with a Hamming-window with filter order q = 64 into characteristic frequency bands

2.2.2 Feature extraction

2.2.2.1 Signal segmentation and feature aggregation

EEG-signal is commonly analyzed in epochs of 30 seconds. Many publications on arousal detection divide the signal in fixed epochs of 20 to 30 seconds and classify if a segment contains an arousal or not. This idea arises from visual scoring and conventional sleep staging. The expert monitors 30s of the EEG signal and decides on the sleep stage and if it contains an arousal by marking it manually. An adaptive segmentation approach as in (Agarwal, 2005), where the segmentation is made due to significant changes in the signal, seems to be more suitable. A disadvantage of adaptive segmentation is clearly that the signal is divided in variable length epochs, which could make it more difficult to compare features among segments. In addition, the classification of equal length segments is more reasonable and reliable.

In this work, a combination of a continuous analysis and segmentation in segments of 3s is realized. The later described classification algorithm has shown the best performance on 3s segments. Features are extracted either per second or per segment. The features extracted per second are aggregated within each segment as further described.

2.2.2.2 Feature extraction derived from the definition of the AASM

This section will be concerned with the extraction of features for the arousal start detection. All features will be derived from the definition of the AASM. Hence only frequency based features will be

extracted without making any use of additional channels except for the available EEG-channels C3A2 and C4A1.

For better understanding the definition of an arousal of the AASM may be repeated here:

"Score arousals during sleep stages N1, N2, N3 or R if there is an abrupt shift of EEG frequency including alpha, theta and/or frequencies greater than 16 Hz (but not spindles) that lasts at least 3 seconds, with at least 10 seconds of stable sleep preceding the change. Scoring of arousal during REM requires a concurrent increase in submental EMG lasting at least 1 second. […]" (Berry, et al., 2016)

The definition gives a relatively clear understanding of the start of an arousal. The duration and the end of an arousal are discussion points and highly depend on the patient, on the sleep expert scoring it and on the appearance of an arousal. In practice, the end of an arousal is set when the signal returns approximately to the state before the arousal.

In this thesis, a two-step detection algorithm will be presented (see Figure 5). In the first step features for the detection of arousal starts are extracted and aggregated in segments. After classifying the segments as arousal start and no-start segments, the exact start and end of the arousals is determined. At the end a check for further arousal criteria, like the minimum duration of 3s, the obligatory 10s of stable sleep preceding an arousal and the increase in submental EMG while REM-sleep, is implemented. With this approach, it will be possible to score with high accuracy the exact position of an arousal.

Features for the detection of arousal starts

To display a shift in spectral content, it is necessary to extract features that measure the signal development and change over time. The point in time where 10s of stable sleep precedes an abrupt shift in frequency of at least 3s describes the start of an arousal. This core idea, taken from (Álvarez-Estévez & Moret-Bonillo, Identification of Electroencephalographic Arousals in Multichannel Sleep Recordings, 2011) is used to analyze the filtered signal. This is done by rolling two consecutive windows, one of 10s and one of 3s, in 1s steps over the whole recording and comparing them with respect to abrupt frequency shifts. The 10s window represents the period of stable sleep and the 3s window a possible arousal start period. The analysis is performed on the filtered signal of both available EEG channels. Figure 13 illustrates the core idea on the example of an arousal displayed on the C3A2 channel. The further described sets of features extracted are measures to describe a difference between these two windows that would indicate an arousal start.

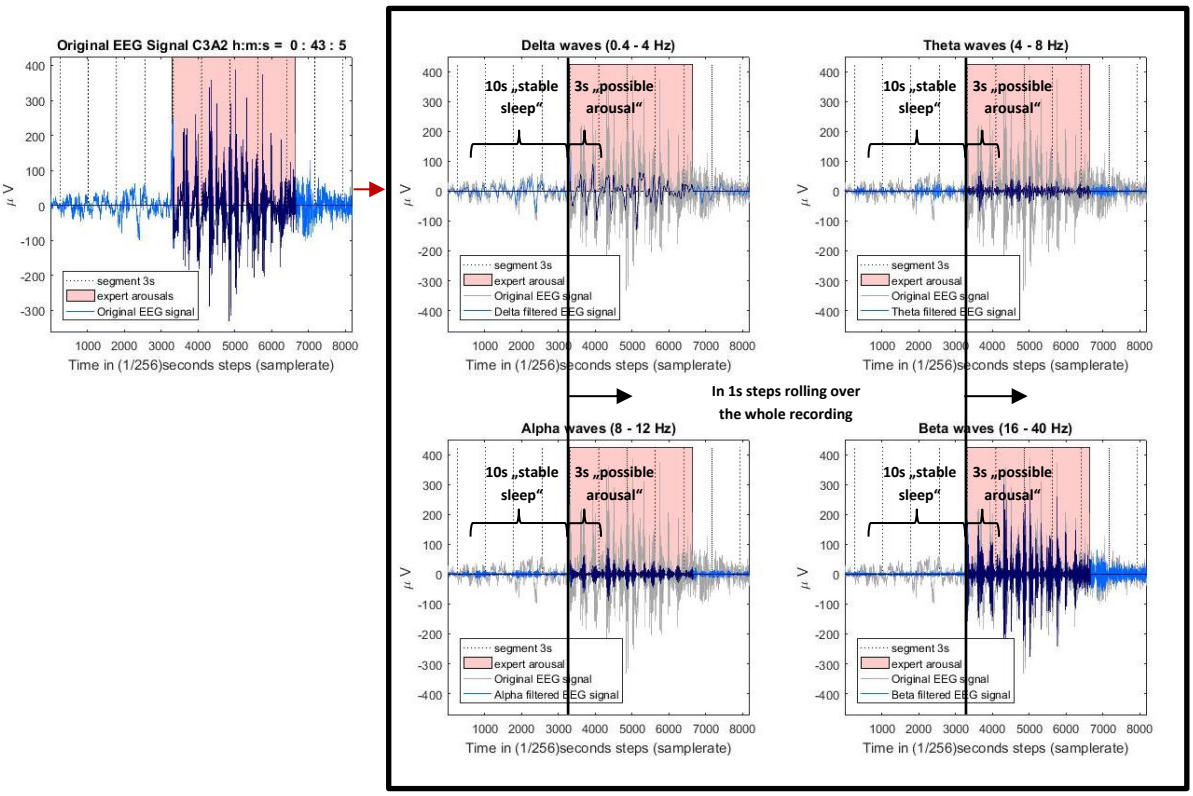


Figure 13: Illustration of the continuous analyzation of the whole recording in 1s steps with two consecutive rolling windows of 10s and 3s

The features are either extracted for each second or within a 3s segment of the recording. Features extracted per second will be summarized in 3s segments. The later described classification algorithms have shown better performance on the 3s segments than on the 1s segments. After the classification, a more precise definition of the arousal start within the classified segments is done by going back to some of the original 1s features. In addition, there are also features extracted without considering the past 10s.

Notation

First, the used notation is explained:

- Let $t = 1, \dots, T$ be the overall time index for all samples of the recording with T being the number of samples (With the considered data, T reaches values between 6 and 8,5 million points, depending on the patient)
- Let $S = 256$ be the sample rate of the EEG-signal
- Let $k = 1, \dots, \frac{T}{S}$ be the index for whole seconds within the recording
- Let $j = 1, \dots, \frac{T}{3S}$ be the index for the segments with a length of 3s

Using these indices, the start of the k^{th} second can be described as: $t(k) = (k-1) \cdot (S) + 1$

Furthermore, certain intervals and segments of the recording need to be described with an adequate notation.

- Let $I_{+n}(k) = \{t \in \mathbb{N} |\ t \in [t(k)\,, t(k+n))\}$ be the interval of all indices t that are part of a n second interval starting in the k^{th} second at $t(k)$ and ending at the start of the second $t(k+n)$ (without the sample at $t(k+n)$).
- Let $I_{-n}(k) = \{t \in \mathbb{N} |\ t \in [t(k-n))\,, t(k))\}$ be the interval of all indices t that are part of a n second interval starting in the $(k-n)^{th}$ second at $t(k-n)$ and ending at the start of the second $t(k)$ (without the sample at $t(k)$).
- For classification the signal is separated in fixed segments of 3s. Using the same notation as for seconds k: 3s segments are notated with: $I_{+3}(3j)$ for $j = 1, \dots, \frac{T}{3S}$

Figure 14 illustrates the used notation for the following derivations.

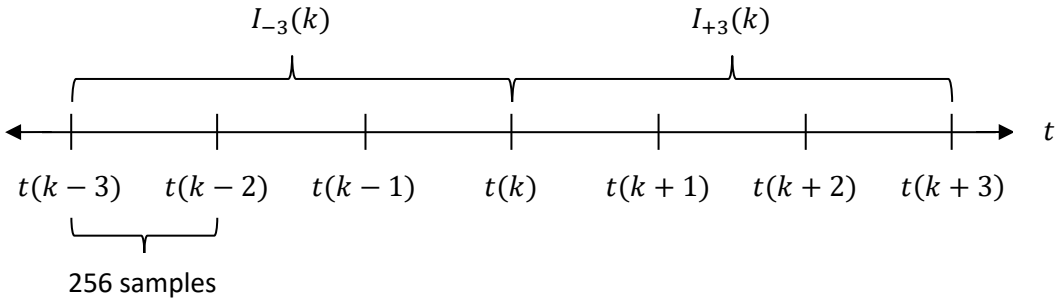


Figure 14: Illustration of notation within one recording

Signal power

The signal power is a possible measure to describe the intensity of a signal within a certain frequency band, which is interesting when searching for frequency shifts. The signal power P within a certain frequency band is defined as the sum of the squared samples of the filtered signal y_t.

Let $y_b(t),\ t = 1, \dots, T\ and\ b = \delta, \theta, \alpha, \beta,$ be the bandpass filtered signal within a characteristic frequency band at time t. And let $X(t)$ describe the whole signal (bandpass filtered within [0.4,40] Hz). The cardinality of a set X is described by $|X|$. Then the total signal power P_b within a certain frequency band b for a certain interval $I_{\pm n}(k)$ is defined as:

$$P_b(I_{\pm n}(k)) = \frac{1}{|I_{\pm n}(k)|} \sum_{t \in I_{\pm n}(k)} y_b^2(t)\,, \qquad b = \delta, \theta, \alpha, \beta$$

The relative signal power is the part of the total signal power that is generated by a certain frequency band b within a certain interval $I_{\pm n}(k)$:

$$p_b(I_{\pm n}(k)) = \frac{\frac{1}{|I_{\pm n}(k)|} \sum_{t \in I_{\pm n}(k)} y_b^2(t)}{\frac{1}{|I_{\pm n}(k)|} \sum_{t \in I_{\pm n}(k)} X^2(t)}, \qquad b = \delta, \theta, \alpha, \beta$$

Feature Sets

The following sets are extracted:

- **Change of total signal power within characteristic frequency bands**
  During an arousal power is shifted from one frequency band to another. Normally power is increasing in total and especially in higher frequency bands like in α- and β-band. By comparing the power of the 10s window before second k and the 3s window after second k the first set of features is calculated (Álvarez-Estévez & Moret-Bonillo, Identification of Electroencephalographic Arousals in Multichannel Sleep Recordings, 2011): Let $P_b(I_{-10}(k))$ be the signal power within a certain frequency band b of the 10s window and $P_b(I_{+3}(k))$ the one of the 3s window.

$$e_b(k) = \frac{P_b(I_{+3}(k))}{P_b(I_{-10}(k))} = \frac{\frac{1}{|I_{+3}(k)|}\sum_{t\in I_{+3}(k)} y_b^2(t)}{\frac{1}{|I_{-10}(k)|}\sum_{t\in I_{-10}(k)} y_b^2(t)}, \qquad b = \delta, \theta, \alpha, \beta \text{ and } k = 1, \dots, \frac{T}{S}$$

  - *4 features per channel* are extracted: $e_b(k)$ is calculated for each second of the recording and for 4 frequency bands.
  - *Aggregation:* For classification the $max(e_b(k))$ within 3s segments is calculated.

- **Change of relative signal power within characteristic frequency bands**
  During an arousal frequencies are shifted to other bands and therefore the change of the part of total power of one frequency band is indicating a shift in frequency. In order to compare if the part of total power that is generated by a certain frequency band is increasing or decreasing, the ratio of relative power $r_b(k)$ between the two windows is calculated:

$$r_b(k) = \frac{p_b(I_{+3}(k))}{p_b(I_{-10}(k))} = \frac{\frac{\frac{1}{|I_{+3}(k)|}\sum_{t\in I_{+3}(k)} y_b^2(t)}{\frac{1}{|I_{+3}(k)|}\sum_{t\in I_{+3}(k)} X^2(t)}}{\frac{\frac{1}{|I_{-10}(k)|}\sum_{t\in I_{-10}(k)} y_b^2(t)}{\frac{1}{|I_{-10}(k)|}\sum_{t\in I_{-10}(k)} X^2(t)}}, \qquad b = \delta, \theta, \alpha, \beta \text{ and } k = 1, \dots, \frac{T}{S}$$

  - *4 features per channel* are extracted: $r_b(k)$ is calculated for each second of the recording and for 4 frequency bands.
  - *Aggregation:* For classification the $max(r_b(k))$ within 3s segments is calculated.

- **Ratio of change in higher frequencies vs. change in lower frequencies**
  A good indicator of a shift in frequency is when the part of power in lower frequencies (θ-band) is decreasing and the part of power in higher frequencies (α- and β-band) at the same time increasing. Thus, by calculating their ratio, a new set of features is extracted:

$$d_{\theta vs(\alpha+\beta)}(k) = \frac{r_\alpha(k) + r_\beta(k)}{r_\theta(k)}, \qquad k = 1, \dots, \frac{T}{S}$$

  - *1 feature per channel* is extracted: $d_{\theta vs(\alpha+\beta)}(k)$ is calculated for each second of the recording.

- *Aggregation:* For classification the $mean(d_{\theta vs(\alpha+\beta)}(k))$ within 3s segments is calculated.

- **Band percentage**
  During an arousal the relative signal power of higher frequencies is greater. The relative signal power within segments of 3s is calculated. (Shahrbabaki, Dissanayaka, Patti, & Cvetkovic, 2015)

$$p_b(I_{+3}(3j)) = \frac{\frac{1}{|I_{+3}(3j)|}\sum_{t\in I_{+3}(3j)} y_b^2(t)}{\frac{1}{|I_{+3}(3j)|}\sum_{t\in I_{+3}(3j)} X^2(t)}, \qquad b = \delta, \theta, \alpha, \beta \text{ and } j = 1, \dots, \frac{T}{3S}$$

  - *4 features per channel* are extracted: $p_b(I(j))$ is calculated for each 3s segment.

- **The center of frequency**
  The next approach is to calculate the center of frequency within one segment, because one may suppose that it shows higher values during an arousal. Therefore a mean value of all possible frequency bands is calculated (Shahrbabaki, Dissanayaka, Patti, & Cvetkovic, 2015):

$$\bar{f}(I_{+3}(3j)) = \frac{\left[P_\delta(I_{+3}(3j))\cdot\left(0.4+\frac{4-0.4}{2}\right)+P_\theta(I_{+3}(3j))\cdot\left(4+\frac{8-4}{2}\right)+P_\alpha(I_{+3}(3j))\cdot\left(8+\frac{12-8}{2}\right)+P_\sigma(I_{+3}(3j))\cdot\left(12+\frac{16-12}{2}\right)+P_\beta(I_{+3}(3j))\cdot\left(16+\frac{40-16}{2}\right)\right]}{P_\delta(I_{+3}(3j))+P_\theta(I_{+3}(3j))+P_\alpha(I_{+3}(3j))+P_\beta(I_{+3}(3j))}$$

  - 1 feature per channel is extracted: $\bar{f}(I_{+3}(3j))$ is calculated for each 3s segment $(j = 1, \dots, \frac{T}{3S})$.

A more statistical point of view leads to the idea of testing on a significant difference in mean of the squared filtered signal between the 10s and the 3s window.

- **Statistical test on structure break between the two consecutive windows**
  Let $y_b(t), t \in I_{-10}(k) \cup I_{+3}(k)$ be the filtered signal in the joined interval of both windows. To analyze if there is a significant difference in mean or a so called structure break between the two windows, a regression model for the squared filter signal is estimated. As mentioned before, this new interval of 13s is rolling in 1s steps over the whole recording. Let $|I_{-10}(k)| = 10 \cdot S = 2560$ be the length of the 10s window, $|I_{+3}(k)| = 3 \cdot S = 768$ the length of the 3s window and $D(t), t \in I_{-10}(k) \cup I_{+3}(k)$ be a dummy variable, which is defined as:

$$D(t) = 0, \qquad if\ t \in I_{-10}(k)$$
$$D(t) = 1, \qquad if\ t \in I_{+3}(k)$$

  Doing so, the dummy variable $D(t)$ is equal to one in the 3s interval and equal to zero in the 10s interval. The mean of a variable can be estimated by performing a regression of the dependent variable on a constant. It is of interest if there is a difference in mean between the squared filtered signal of the two intervals.

The regression model can be formulated as follows:

$$y_b^2(t) = \beta_{b,0}(k) + \beta_{b,1}(k) \cdot D(t) + u_b(t),$$

$$t \in I_{-10}(k) \cup I_{+3}(k), \qquad b = \delta, \theta, \alpha, \beta, \qquad k = 1, \dots, \frac{T}{S}$$

With $\beta_{b,0}(k)$ and $\beta_{b,1}(k)$ representing the unknown parameters and $u_b(t)$ the errors.

The parameters are estimated with ordinary least squares method, by minimizing the sum of squared errors. $\beta_{b,1}(k)$ can be interpreted as the change in the mean value of y_b^2 if $t \in I_{+3}(k)$, or, in other words, it is the difference in the estimated mean of y_b^2 between the 10s and the 3s window. If $\beta_{b,1}(k)$ is significantly different from zero, the mean in both windows is statistically not the same and one can extract the related t-test statistic as a feature and indicator for a frequency shift. The hypothesis is:

$$H_0: \beta_{b,1}(k) = 0$$
$$H_1: \beta_{b,1}(k) \neq 0$$

Assumed that the above defined regression model fulfills the assumptions of a classical regression model, the t-test statistic $\tau_b(k) = \frac{\widehat{\beta_{b,1}}(k)}{\sqrt{\widehat{Var}(\widehat{\beta_{b,1}}(k))}}$ is under the H_0 and the assumption that the errors u_b are normally distributed

$$u_b \sim N(0, \sigma^2 I_{|I_{-10}(k) \cup I_{+3}(k)|})$$

t-distributed with $|I_{-10}(k) \cup I_{+3}(k)| - 1 = 3327$ degrees of freedom.

Although we can't assume a normal distribution of the errors, we still can calculate the t-test statistic for $\beta_{b,1}(k)$ and use it as a feature set.

$$\tau_b(k) = \frac{\widehat{\beta_{b,1}}(k)}{\sqrt{\widehat{Var}\left(\widehat{\beta_{b,1}}(k)\right)}}, \qquad b = \delta, \theta, \alpha, \beta \text{ and } k = 1, \dots, \frac{T}{S}$$

- *4 features per channel* are extracted: $\tau_b(k)$ is calculated for each second of the recording and for 4 frequency bands.
- *Aggregation:* For classification the $max(\tau_b(k))$ within 3s segments is calculated.

Thinking in terms of time series analysis, one can try to fit an AR-model to the data within the segments and compare the coefficients of the model within arousal segments and no-arousal segments.

- **AR coefficients**
  In each 3s segment (quasi stationary segment) an AR(p)-Model with parameters $a_1, \dots, a_p$ is fitted to the EEG data X_t:

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \cdots + a_p X_{t-p} + u_t$$

- p *features per channel* are extracted: $a_1, \dots, a_p$ is estimated within each 3s segment.

2.2.2.3 *Feature analysis and reduction*

The aim of this section is to analyze which features are relevant for the classification afterwards. Too many features and especially features that have no meaning for the classification can result in a poor performance of the algorithm. It has to be noted that the purpose here is to find indicative features for arousal starts. The duration for detected arousal is calculated in a second step.

Before analyzing the relevance of the features for the classification, a short analysis is done on the AR-Model features to determine a model order.

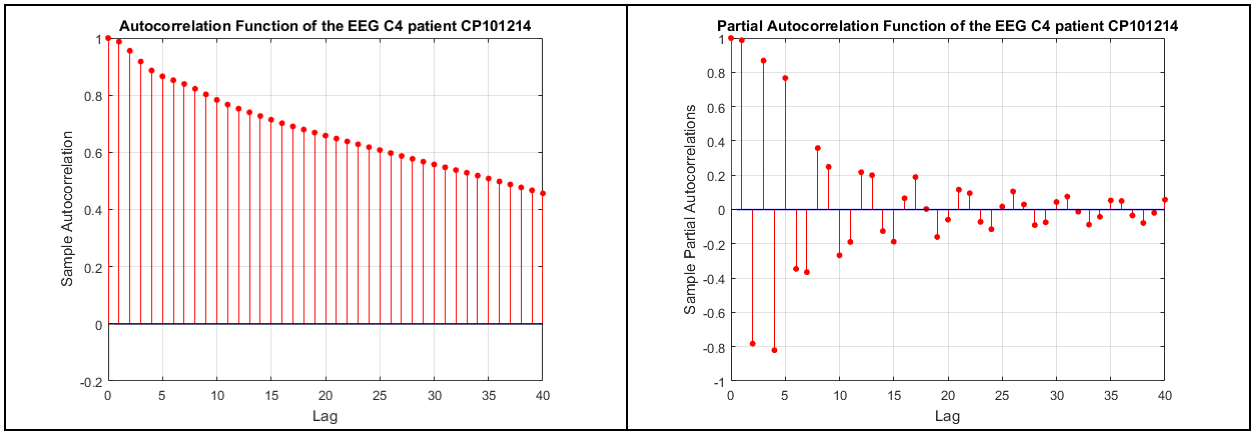
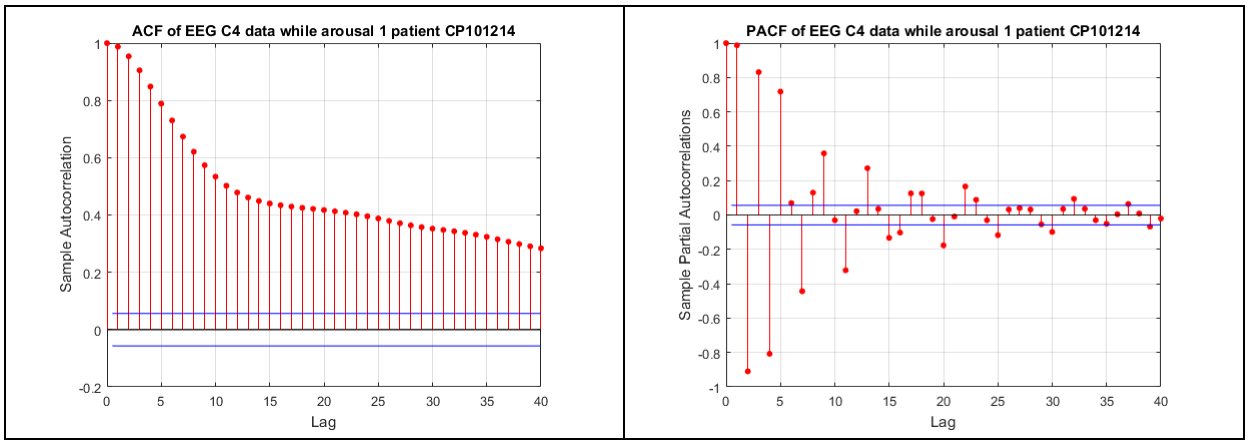


Figure 15: Autocorrelation function and Partial Autocorrelation function for the EEG C4A1 signal of patient CP101214

Figure 15 shows the ACF and the PACF of the whole recording of the EEG channel C4A1 of patient CP101214. Due to the high amount of data points the correlation and the partial correlation for all lags are significant. A simple and flexible way to model this, is using an AR-model. There are publications on this topic, for example (Zhang, Ji, & Zhang, 2015) where more can be read about classification of EEG-signal with AR-models. In this work the coefficients are suggested as features and therefore no further time series analysis is done.

Figure 16 shows how ACF and PACF differ between two randomly chosen segments, one with and one without an arousal.



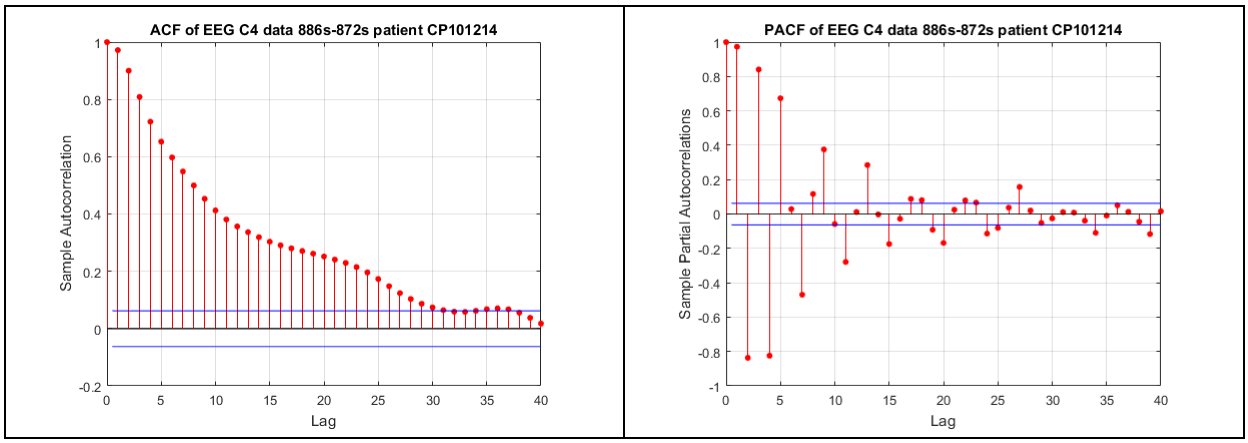


Figure 16: First row shows the ACF and the PACF during a segment of the first arousal of patient CP101214 and the second row shows the same for a random chosen no-arousal segment

Features will be extracted within segments of 3s and therefore the order p of the AR(p)-Model for X_t within one segment is estimated with BIC information criteria for the whole recording.

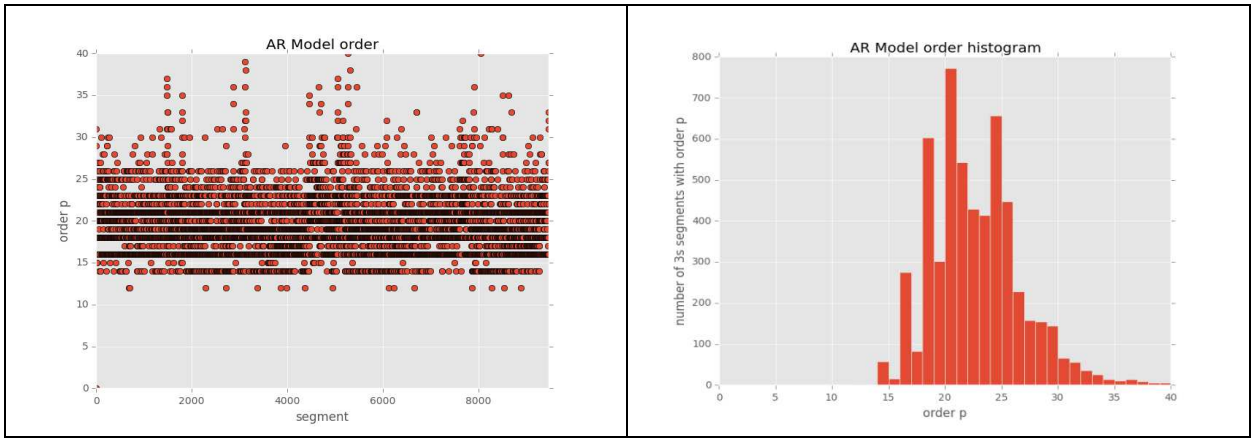


Figure 17: Distribution of AR model order p estimated with BIC criterion for the C3A2 data of patient TS100714

Figure 17 shows the distribution of the model order within one segment along the recording of one patient (other patients show similar results). The order $p = 20$ seems to be a good approximation for the estimated order. The order is fixed at $p = 20$ and development of the parameters (estimated within each segment) for the whole recording is plotted for patient MP020714 in Figure 18 (lags 1-10) and in Figure 19 (lags 11-20). In addition, the center of frequency within each segment is plotted below and arousal segments are marked with black lines.

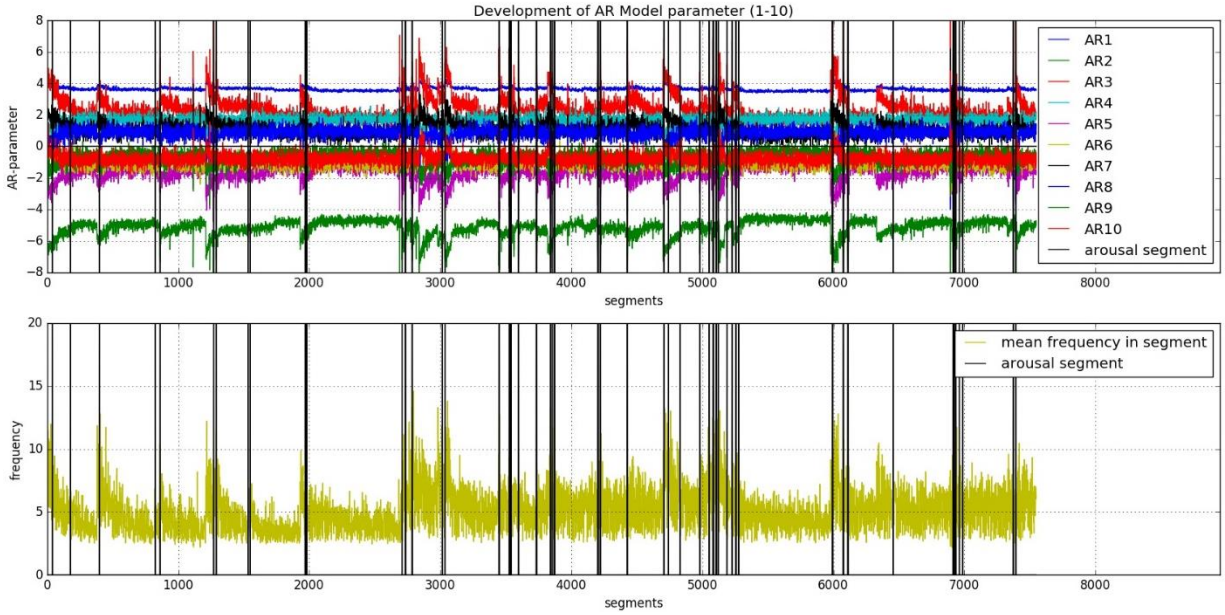


Figure 18: Development of AR(20) parameters for lags 1-10 and the center of frequency for the same segments; Patient: MP020714; EEG-channel: C3A2

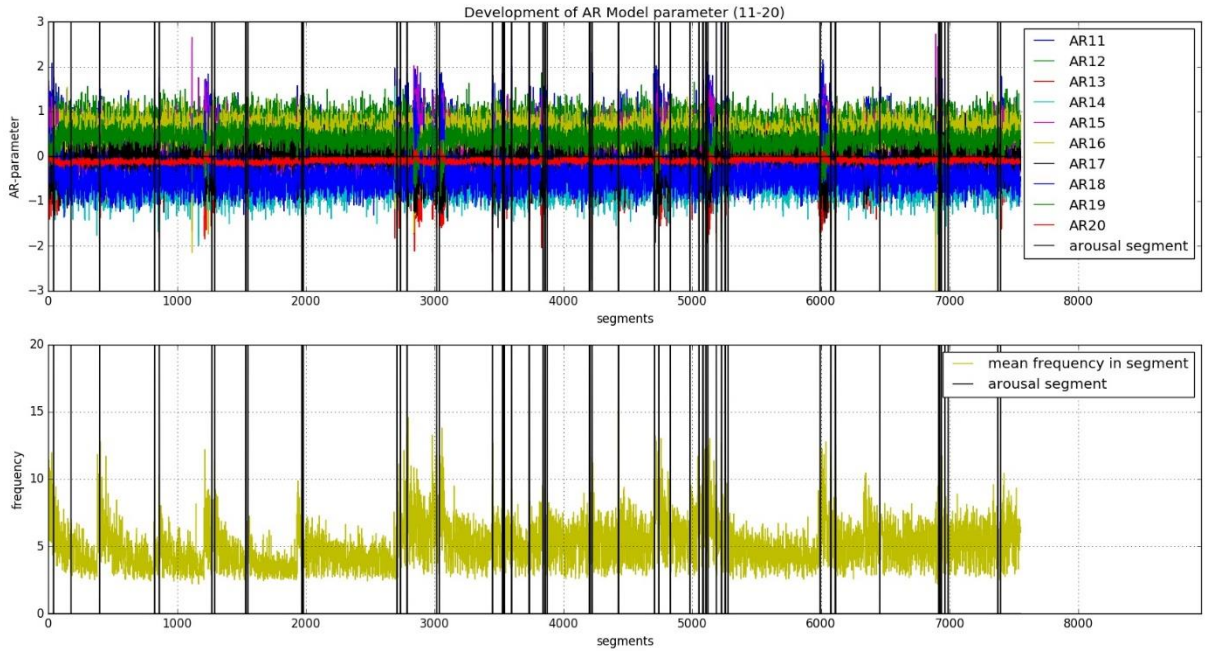


Figure 19: Development of AR(20)-parameter for lags 11-20 and the center of frequency for the same segments. Patient: MP020714; EEG-channel: C3A2

Furthermore it seems that the coefficients of the AR(20)-model are quite stable with the exception of time intervals close to arousals. It can be observed that especially the coefficients for smaller lags are more sensible to segments, in which the center of frequency is increasing and arousals are occurring. Although the estimation of the order of the model resulted in order $p = 20$, it appears that smaller lags may contain more information about arousals (see boxplots in Figure 23) and therefore a smaller model with order $p = 6$ is tested for feature extraction (see boxplots in Figure 24).

Figure 20 shows that the coefficients of the AR(6)-model, with exception of the coefficients for lag 1 and lag 6, react sensible in segments where the center of frequency is increasing and arousals are occurring. This characteristic is useful for the detection of arousals.

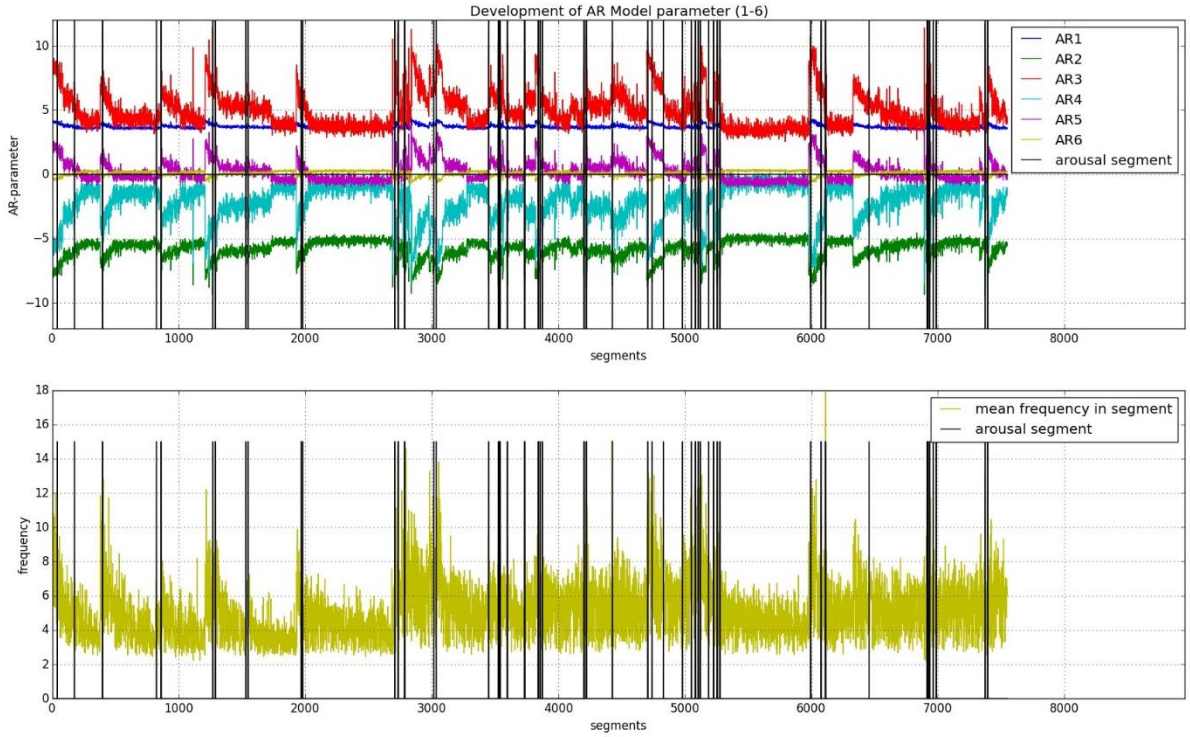


Figure 20: Development of AR(6)-parameter for lags 1-6 and the center of frequency for the same segments. Patient: MP020714: C3A2

In order to examine in greater detail how features are distributed between no-arousal start segments (0) and arousal start segments (1) one possibility is to visualize them using boxplots. Figure 21 and Figure 22 show boxplots of all calculated features except for the AR(p)-coefficients.

The high amount of outliers can have different reasons. One reason is that EEG-signal is in general an unstable data set and there are many artifacts. In this work, there is no removal of artifacts performed when preprocessing data and therefore signal reaches values high above the average in moments of amplitude artifacts and causes outliers when calculating features, that use squared values of the signal. In most of the cases the boxes in the boxplot overlap between the arousal start segment group and no arousal start segment group. This can be partly caused by facts like, that arousals may have very different appearances (see Figure 4), that also periods without arousals can show abrupt frequency shifts and that segments directly next to arousal starts can show similar values of features but are labeled as no-arousal start segments.

In Figure 21 and Figure 22 features in a red box show a clearer difference between arousal and no arousal segments than others and are chosen for classification process. Features calculated exclusively with delta filtered signal aren't considered for the classification process because according to the AASM definition of an arousal, a shift in delta frequency is not counted as an arousal. In addition, delta artifacts are quite common in EEG-signal and the inclusion of delta features would cause wrong detection of artifacts instead of real arousals.

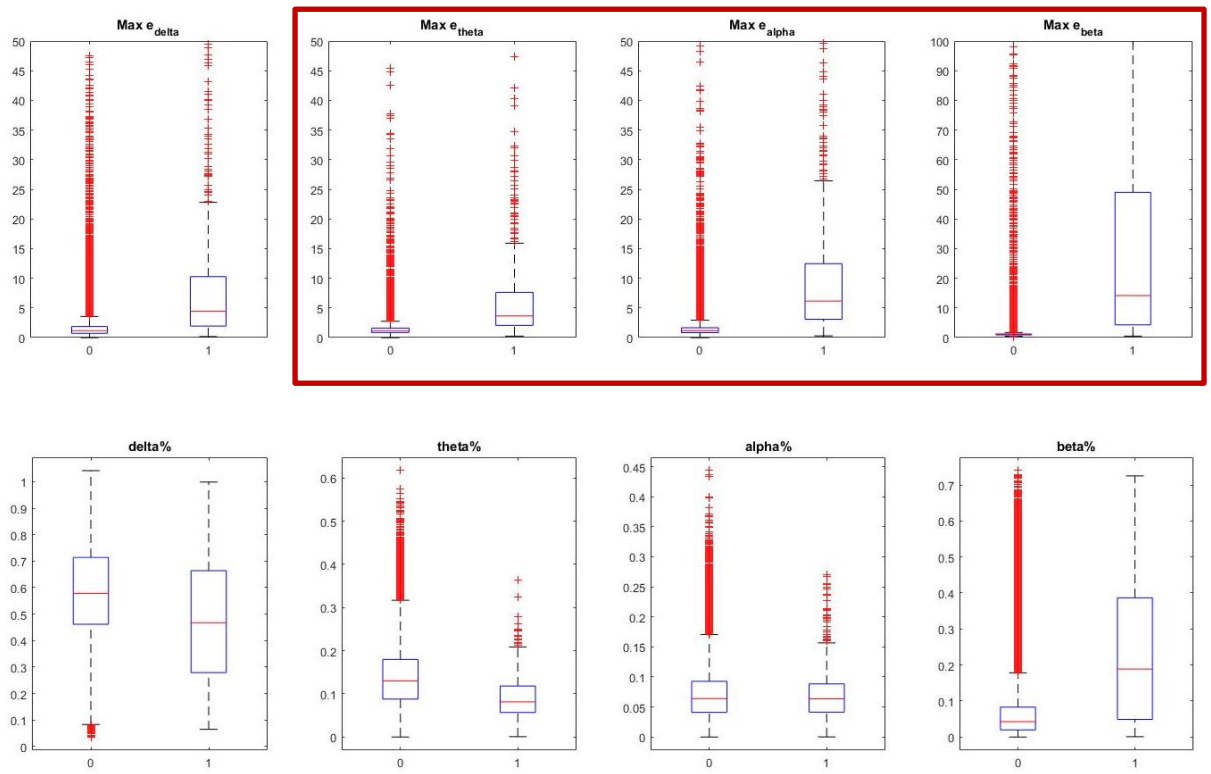


Figure 21: Boxplots of 8 features between arousal start segments (1) and no arousal start segments (0) of 3s of channel C3A2.

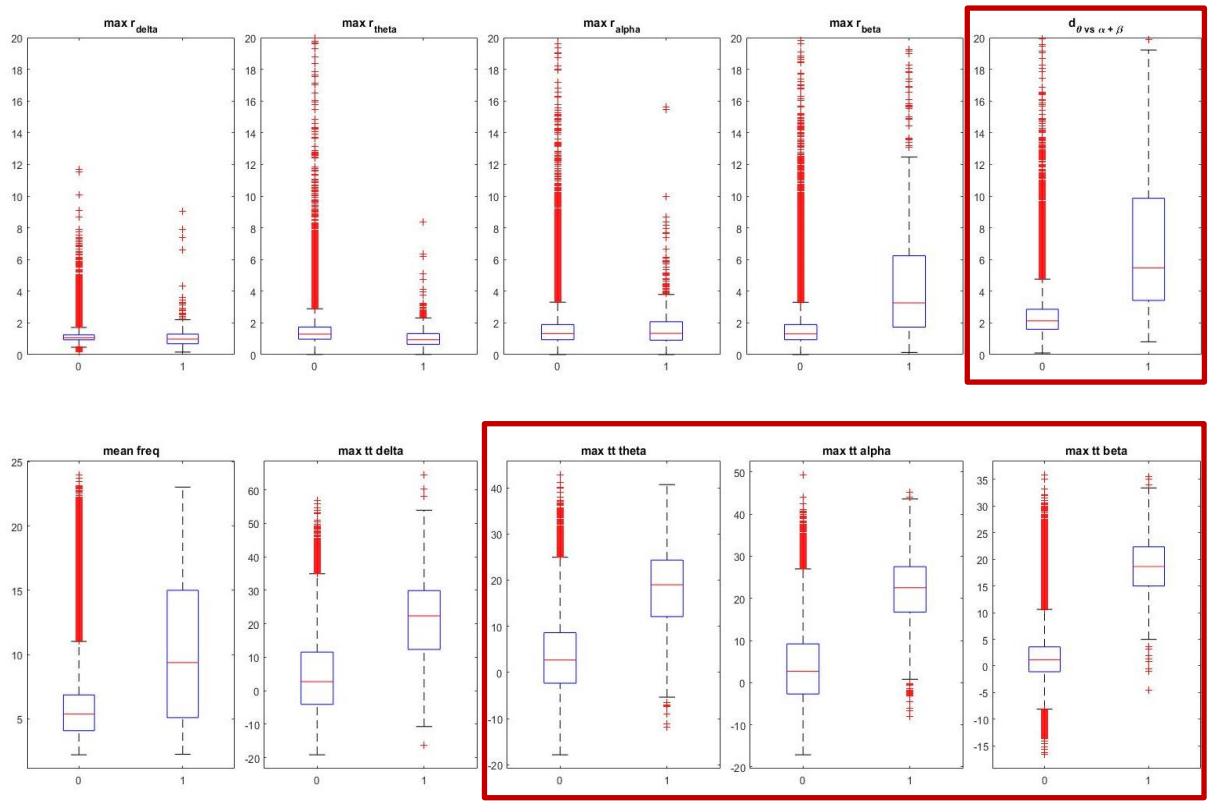


Figure 22: Boxplots of 10 features between arousal start segments (1) and no arousal start segments (0) of 3s of channel C3A2.

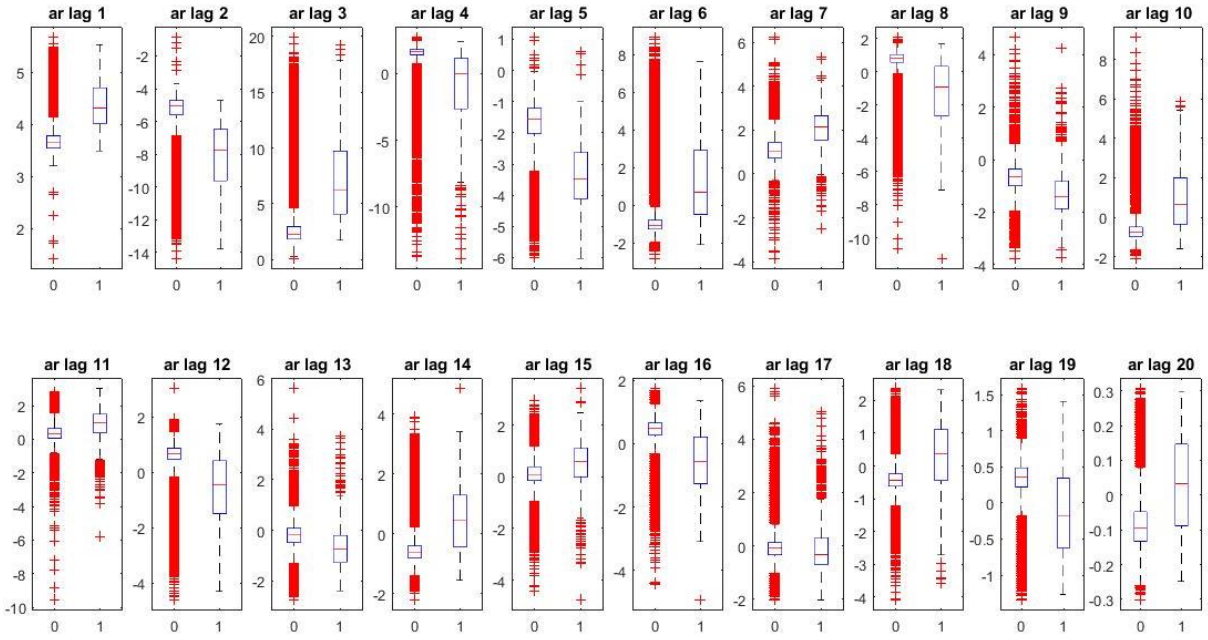


Figure 23: Boxplots of AR(20)-coefficients until max lag = 20 between arousal start segments (1) and no-arousal start segments (0) of 3s

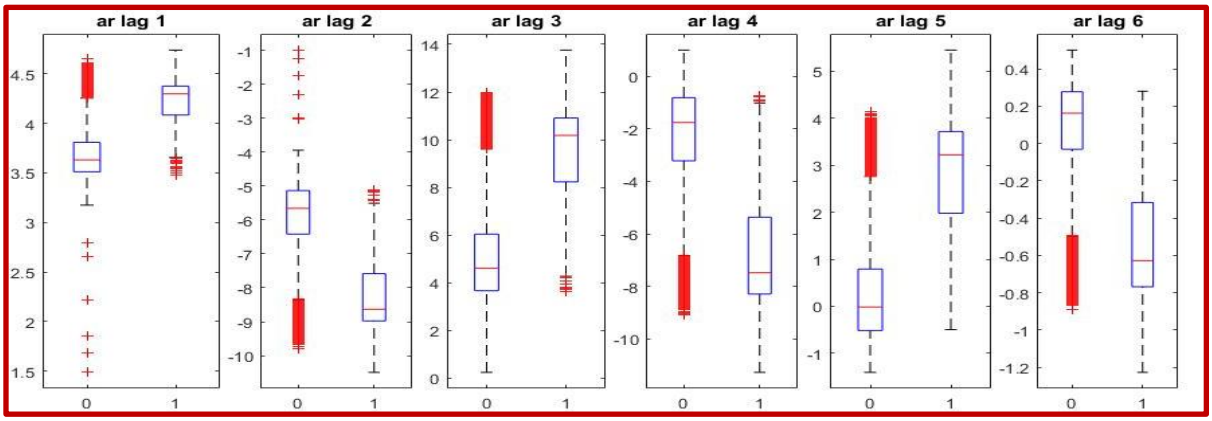


Figure 24: Boxplots of AR(6)-coefficients until max lag = 6 between arousal start segments (1) and no-arousal start segments (0) of 3s

Figure 23 shows the boxplots of the coefficients of the estimated AR(20) model for each segment. Compared to Figure 24, which provides the boxplots of the coefficients of the AR(6) model for each segment, it can be seen that the boxes never overlap for the AR(6) model and a clearer difference in distribution between arousal and no-arousal start segments can be seen. Although the order $p = 6$ may not be the real order of the model, the coefficients of this model are included in the classification algorithm.

The performance of the later described classification algorithm was tested with backward feature reduction, by successively removing features and staying with the smallest set of features (marked with red boxes in Figure 21 and Figure 22 plus the coefficients of the AR(6) model) that gave the best performance results. In the end, 13 features times 2 channels (26 features) were chosen for the classification process.

2.2.3 Preparation for Classification

2.2.3.1 Definition of arousal start segments

For the training of the classifier, it has to be decided which segments are considered to be true arousal start segments. As segmentation is done strictly in 3s segments, it can happen that the start of the annotated arousal is just at the very end of a segment. In this case the next segment should be considered as the real start segment to avoid "bias" in the level of features within real start segments. Arousal segments are labeled according to the following condition:

- A segment is labeled to be an arousal segment if at least 15% of the 3s segment is part of the arousal.
- The first segment of each group of 3s segments describing an arousal is marked as the arousal start segment, which is used for classification.

Figure 25 shows on the left side the case in which more than 15% of the start segment are part of the arousal, and on the right side the case in which less than 15% of the start segment are part of the arousal and the next segment is considered to be the real start segment.

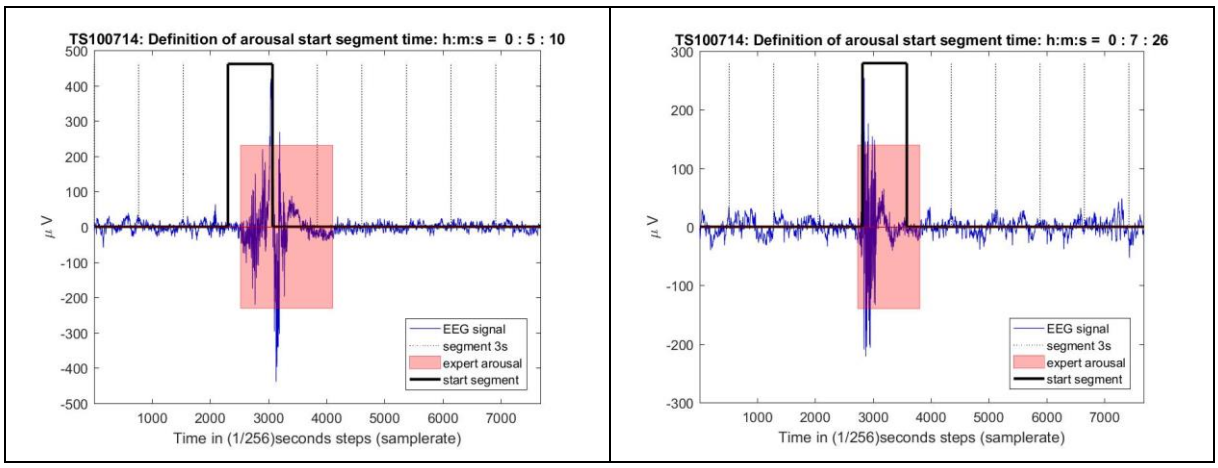


Figure 25: Illustration of the definition of an arousal start segment. Patient TS100714; EEG-channel: C4A1

2.2.3.2 Check of hypnogram

Referring to Note 3 and Note 5 of the AASM definition of an arousal, it is possible to score arousals during awakening. But from a clinical point of view and according to the practice of manual scoring, it is not common to score arousals during a phase where the patient is awake. Hence, all segments where the hypnogram appears to be zero in any moment are excluded from further arousal detection.

Another critical point is how to distinguish between a real arousal and a transition to wake. In practice, there is either an arousal or a transition to wake scored. Therefore, due to the fact that the sleep stages are given and trusted, the last 30s (the last 10 segments) before hypnogram turns zero are also excluded.

Referring to Note 4, the 10s of stable sleep can start during an epoch staged as wake. Therefore, nothing will be excluded after an epoch with hypnogram equal to zero.

2.2.4 Classification

There are many different options to classify segments. In literature, machine learning algorithms as well as different discriminant functions (Álvarez-Estévez & Moret-Bonillo, Identification of Electroencephalographic Arousals in Multichannel Sleep Recordings, 2011), K-Nearest-Neighbors classification (Shahrbabaki, Dissanayaka, Patti, & Cvetkovic, 2015) or simple threshold decision algorithms (Agarwal, 2005) are applied. A great number of publications on this topic rely on machine learning algorithms because of better performance as already researched for example in (Álvarez-Estévez & Moret-Bonillo, Identification of Electroencephalographic Arousals in Multichannel Sleep Recordings, 2011). Thus, in this thesis a support vector machine (SVM) is applied.

For validation of the detection performance tested on the 7 patients of AKH Wien a "Leave-One-Out-Cross validation" is used. That means that 6 patients are used as training set and the remaining one patient is used to validate the algorithm. Here the python support vector machine algorithm is used out of the `scikit-learn` module. (scikit-learn, 2010-2016)

2.2.4.1 Support Vector Machine

The aim of this section is to give a short overview of the functionality of a SVM. Due to the fact, that the algorithms were implemented in **python**, the following methods are derived based on the book *Introduction to Machine Learning with Python* (Guido & Müller, 2016) and the documentation of the *scikit-learn* package (scikit-learn, 2010-2016).

A SVM constructs a hyper-plane (in 2D a line) or set of hyper-planes in a high or infinite dimensional space, which can be used for classification, regression or other tasks. Intuitively, a good separation is achieved by the hyper-plane that has the largest distance to the nearest training data points of any class (so-called functional margin), since in general, the larger the margin, the lower the generalization error of the classifier. (scikit-learn, 2010-2016)

Support vectors are the feature-vectors within the training set that lie closest to the separation hyper-plane and are therefore the most difficult to classify. A support vector machine finds the optimal hyper-plane to separate the dataset by maximizing the margin between them (see Figure 26).

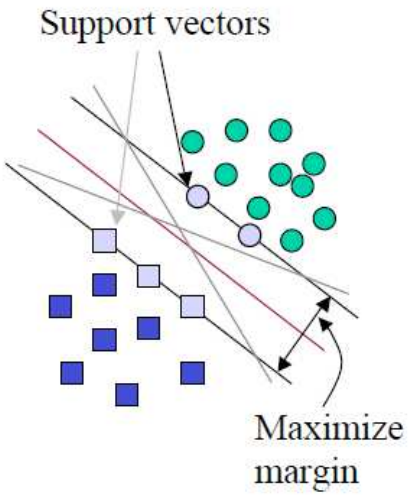


Figure 26: Illustration of the maximization of the margin between the classes done by the SVM (scikit-learn, 2010-2016)

A SVM solves an optimization problem for the weights that correspond to the support vectors. Compared to linear regression or neural nets the important difference in support vector machines is that only some points and not all of them are considered when finding the optimal separation.

SVM with kernel function

If there is no linear hyperplane that separates the groups, the feature space can be mapped into a higher dimensional space to find a linear separation. This can be realized by substituting the inner product $x_i \bullet x_j$ in the optimization problem *(4)* with a kernel function $K(x_i, x_j) = \varphi(x_i)'\varphi(x_j)$. The function φ transforms the features into a higher dimensional space and is defined implicitly by the kernel function K. Figure 27 shows how the separation hyperplanes between the classes are changing when using different kernel functions.

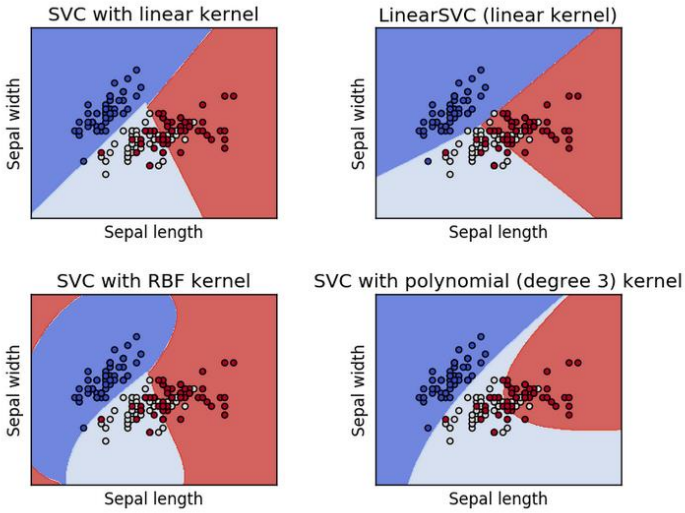


Figure 27: Illustration of different kernel functions for SVM (scikit-learn, 2010-2016)

SVM optimization in $\mathbb{R}^2$

Figure 28 shows an example for a 2D-space with a linear hyperplane to separate the training set. Let $x_i \in \mathbb{R}^2$, $i = 1, \ldots, N$ be the training vectors in two classes and $y \in \{-1,1\}^N$ a vector to label the classes, and $\|x\|$ be the Euclidian norm of x.

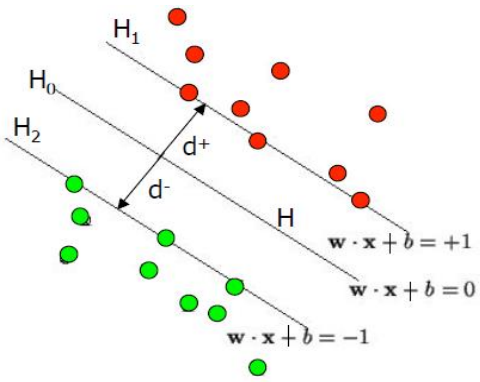


Figure 28: Illustration of SVM optimization (scikit-learn, 2010-2016)

The hyperplanes H_1 and H_2 can be defined as:

- H_1: $x_i \bullet \omega + b = +1$
- H_2: $x_i \bullet \omega + b = -1$

The distance between a point x and a line $\{y \mid y \bullet \omega + c = 0\}$ can be calculated as:

$$distance(x, \{y \mid y \bullet \omega + c = 0\}) = \frac{|x \bullet \omega + c|}{\|\omega\|}$$

Since it holds for all points on H_1 that $|\omega \bullet x + b| = 1$, the distance d between H_0 and H_1 is given by:

$$d = \frac{|\omega \bullet x + b|}{\|\omega\|} = \frac{1}{\|\omega\|}$$

Hence, the total distance between H_1 and H_2 is $\frac{2}{\|\omega\|}$.

To maximize the distance, we need to minimize $\|\omega\|$ under the condition that there are no data points between the hyper-planes. This condition can be formulated as:

$$x_i \bullet \omega + b \geq +1 \; when \; y_i = +1$$
$$x_i \bullet \omega + b < -1 \; when \; y_i = -1$$

By combining them into one condition we obtain: $y_i(x_i \bullet \omega + b) \geq 1$

SVM optimization in $\mathbb{R}^p$

Given training vectors $x_i \in \mathbb{R}^p$, $i = 1, \ldots, N$ in two classes and a vector $y \in \{-1,1\}^N$. Let ω be the vector of weights corresponding to the support vectors. C is a constant to scale the penalty term and ζ_i are the so called slack variables to soften the condition.

A SVM solves the following primal problem:

$$\min_{\omega,b,\zeta} \frac{1}{2}\omega'\omega + C\sum_{i=1}^{n}\zeta_i$$
$$s.t. \quad y_i(\omega'\varphi(x_i) + b) \geq 1 - \zeta_i \qquad (3)$$
$$\zeta_i \geq 0, i = 1, \ldots, n$$

The function φ is implicitly defined by the kernel function $K(x_i, x_j) = \varphi(x_i)'\varphi(x_j)$, and maps the features into higher dimensional space, when no linear separation can be found directly. In the study of (Álvarez-Estévez & Moret-Bonillo, Identification of Electroencephalographic Arousals in Multichannel Sleep Recordings, 2011) as well as in other studies, that are making use of SVM for classification, a SVM with radial basis function for the kernel is used. A radial basis function is a kernel function defined as $K(x, y) = \exp(-\gamma\|x - y\|^2)$ with $\gamma = \frac{1}{2\sigma^2}$.

The dual problem of the above formulated primal optimization problem can be defined as follows:

$$\min_{\alpha} \frac{1}{2}\alpha' Q\alpha - e'\alpha$$
$$s.t. \quad y'\alpha = 0 \qquad (4)$$
$$0 \leq \alpha_i \leq C, i = 1, \dots, n$$

Where e is the vector of ones, $C > 0$ is the upper bound, Q is an nxn positive semidefinite matrix, $Q_{ij} = y_i y_j K(x_i, x_j)$ where $K(x_i, x_j) = \varphi(x_i)'\varphi(x_j)$ is the kernel. (scikit-learn, 2010-2016)

The decision function can be formulated as:

$$sgn(\sum_{i=1}^{n} y_i \alpha_i K(x_i, x) + \rho)$$

A critical point when using SVMs is the selection of the parameters C and γ. Intuitively, the parameter γ describes how far the influence of a single training value reaches, low values of gamma meaning far and high values of γ meaning close. The parameter C deals with misclassification of training samples. The complexity of the decision surface is highly determined by the selection of the parameter C. High values of C mean that the algorithm tends to not misclassify any trainings set by choosing more support vectors and providing a rather complex decision surface. If the parameter C is chosen low the simplicity of the decision surface has more importance than the possible misclassification of trainings sets.

In general, the behavior of the SVM is more sensitive to the parameter γ than to the parameter C. On the one hand, if γ is chosen too large, the region of influence of one support vector included only itself and no more regulation can be done by modifying C. On the other hand, if γ is chosen too small, the model is similar to a linear model and the complexity can be increased by choosing higher values of C, which gives the model the freedom of selecting more support vectors. Figure 29 is a visualization of

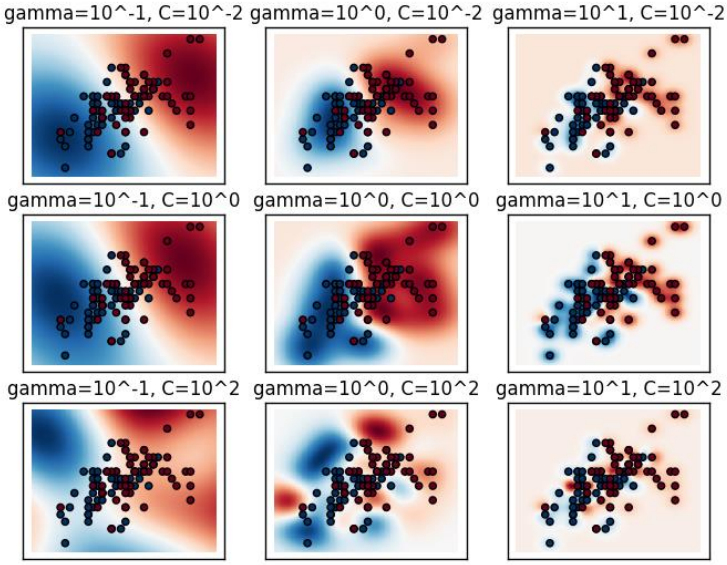


Figure 29: Illustration of influence of different choices of parameters C and gamma on the decision function (scikit-learn, 2010-2016)

the decision function for different choices of parameters C and γ of a classification problem with only two features and two possible classes.

The arousal detection problem is a highly unbalanced one. In fact, there are a lot more segments not containing arousals than actual arousal segments. This has to be considered in the classification by choosing class weights.

2.2.4.2 Validation of classifier

In this work, it has to be differentiated between the validation of the classifier, to measure its performance and to choose the best parameters, and the validation of the overall detection algorithm. In the case of the validation of the classifier there is a clear true classification decision on each 3s segment if it is an arousal start segment or not. Each arousal has one unique start segment (as explained in section *determination of arousal start segments*) that is desired to be detected by the classifier.

The following common statistical terms are used to calculate how good the classifier performs in detecting arousal start segments:

- TP (True Positives) = Number of events detected which matched with the actual events
- FP (False Positives) = Number of events detected which did not match with the actual ones
- TN (True Negatives) = Number of no-events which matched correctly
- FN (False Negatives) = Number of no-events which did not match correctly

In the case of arousal start segment classification it is clear how to interpret these terms:

- TP (True Positives) = Number of annotated arousal start segments that are correctly detected
- FP (False Positives) = Number of detected arousal start segments that don't match with the actual annotated segments
- FN (False Negatives) = Number of annotated arousal start segments that are not detected
- TN (True Negatives) = Number of correctly detected no-start segments

As the classifier makes decisions without knowledge of the neighboring segments, a detection of a neighboring segment is also counted as a False Positive. Trials on detection of start areas, in terms of detected neighboring segments performed worse.

The following statistical values are considered for validation:

- *True Positive Rate (Sensitivity):* $\frac{TP}{TP+FN}$
  (Percentage of all true arousal starts that were correctly detected - capability of the process to detect arousal starts*)*
- *Specificity:* $\frac{TN}{TN+FP}$
  (Percentage of all true no-starts that were correctly detected - capability of the process to detect that there is no arousal start*)*
- *False Negative Rate:* $\frac{FN}{TP+FN} = 1 - sensitivity$
  (Percentage of all true arousal starts that were not correctly detected)
- *False Positive Rate:* $\frac{FP}{TN+FP}$

(Percentage of all true no-starts that were not correctly detected)

- *Positive predictive value:* $\frac{TP}{TP+FP}$
  (Percentage of all detected arousal starts that are true arousal starts)
- *Negative predictive value:* $\frac{TN}{TN+FN}$
  (Percentage of all detected no-starts that are true no-starts)
- *Accuracy:* $\frac{TP+TN}{TP+FP+TN+FN}$

2.2.4.3 Parameter search

To choose the optimal parameters for the SVM a grid search is performed. The searched parameter sets are:

- $C \in \{1, \dots, 20, 30, 100\}$
- $\gamma \in \{2^{-7}, 2^{-8}, 2^{-9}, 2^{-10}, 2^{-11}, 2^{-12}, 2^{-13}\}$.

The performance of the classification algorithm is measured with common statistical values described in the section *Validation of classifier.* In order to choose the best parameters for a classifier one has to consider the trade-off between a high sensitivity and the least wrong detections possible.

A common calculation method for the decision of the best parameters is to determine the parameter combination with the maximum of the Youden-Index:

$$Youden\ Index = Specificity + Sensitivity - 1$$

Graphically, this can be displayed by plotting the Receiver Operating Characteristics (ROC) curve (see Figure 30) for all parameter combinations $(C, gamma)$ the Sensitivity (y-axis) is plotted against the False-Positive-Rate (x-axis). The parameters of the point with the maximum distance to the diagonal give the best trade-off between True-Positive-Rate and False-Positive-Rate. (medistat, 2016)

The results for each parameter combination are computed by applying Leave-One-Out-Cross validation and calculating the mean value over all 7 patients. This means that results for each patient are received by training the SVM on the remaining 6 patients and testing it on one patient.

The best parameters according to the Youden-Index and the ROC curve are:

$$C = 11\ and\ \gamma = 2^{-12}\ with\ Youden - Index = 92{,}67\%$$

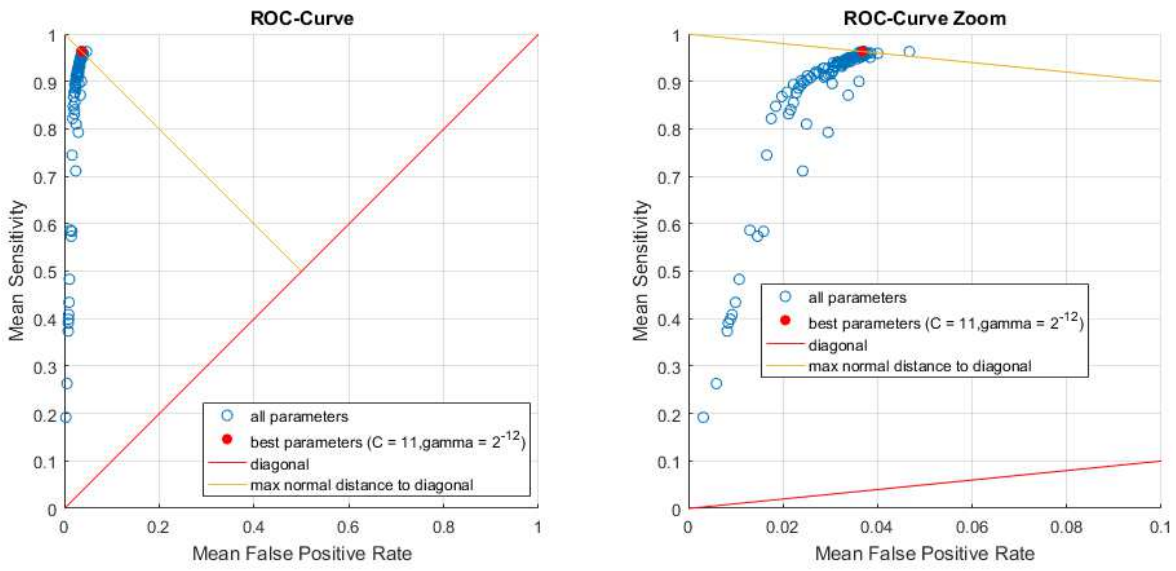


Figure 30: ROC curve of the average of all patients: Sensitivity and False-Positive-Rate for all parameter combinations with marked max distance point to diagonal

2.2.5 Final arousal determination

2.2.5.1 Specification of the arousal start

After detecting arousal starts with the classification algorithm, the detected segments are analyzed and the precise start is set according to the following rules:

- Segments that are are directly next to each other and classified as arousal start segments are joined and counted as one arousal start area. (see yellow area in Figure 31)
- Within one arousal start area it is searched for the maximum of the sum of t-test statistics $\tau_{b,c}(k)$ for $b = \theta, \alpha, \beta$ and summed up for all channels c. This feature set turns out to react sensible and is therefore suitable for precise specification of the start.

$$k^* = \max_{k=1\ldots(\#joined\ segments)\cdot 3} \left(\sum_{c\in\{C3A2,C4A1\}} \tau_{\theta,c}(k) + \tau_{\alpha,c}(k) + \tau_{\beta,c}(k) \right)$$

- The exact start second k^* is set at the maximum.

Figure 31 illustrates the specification of the exact arousal start (maximum of red curve) within the detected start area (yellow area).

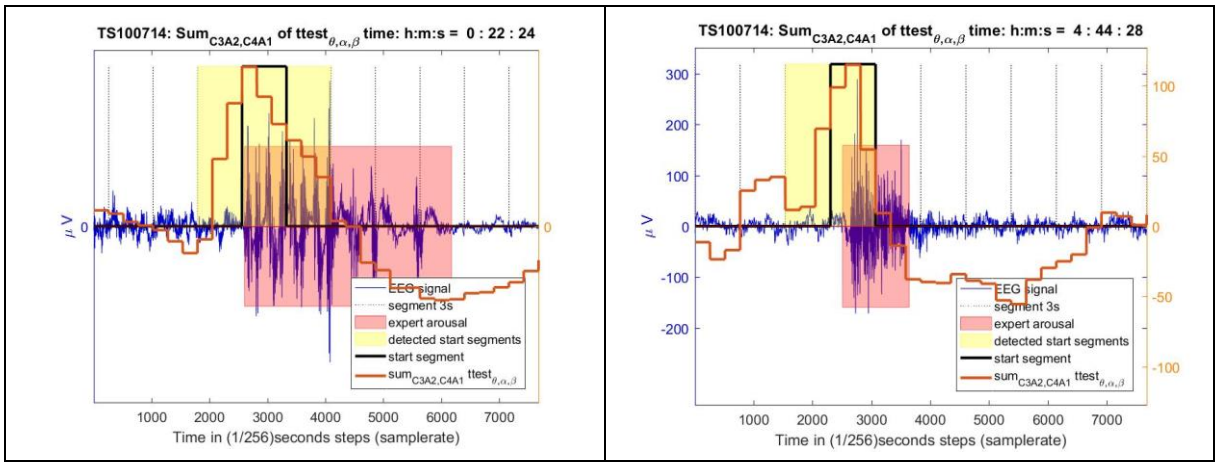


Figure 31: Illustration of specification of the exact arousal start within the detected start area. Start is set where the red curve within the yellow area reaches its maximum. Patient TS100714; EEG-channel: C4A1

2.2.5.2 Calculation of arousal duration

In a next step the duration of the detected arousals needs to be calculated. Therefore, an adjusted version of the core analyzation technique of the 10s stable sleep followed by 3s possible change is applied, because the practice says that the signal needs to return to its state before the arousal. That`s why it seems to be suitable to compare stepwise small intervals after a detected arousal start with its preceding 10s “normal sleep”. Figure 32 should illustrate the following main idea: The 10s before a possible arousal start describe the “normal sleep pattern” to which the EEG-signal should return after the arousal. So this interval needs to be described by a feature that saves the information about how the signal looked like before the start second k^*. The same feature is calculated for a rolling window of 2s (moving on in $0.25s$ steps) after the start and is successively compared (by calculating its ratio) to the 10s interval before the start.

As long as the feature within a 2s interval is significantly greater ($threshold_{duration}$ times greater) than the one while stable sleep, arousal end is not marked yet.

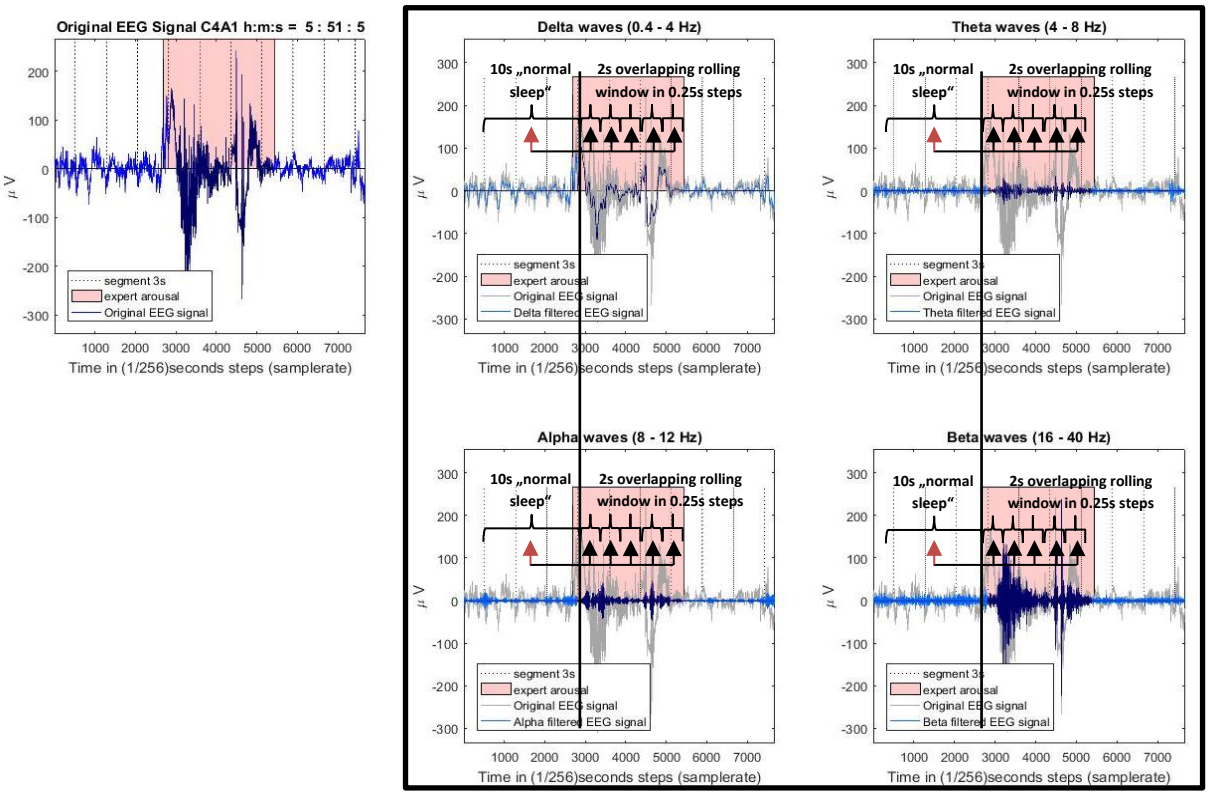


Figure 32: Illustration of the calculation of the arousal duration

A reliable and simple value to compare the 10s normal sleep with a rolling window of 2s (moving on in $0.25s$ steps) is the signal power. Therefore, it is tested if the ratio of the sum of total signal power of the frequency bands θ, α and β within a 2s interval and the signal power within the 10s before the start of an arousal is greater than $threshold_{duration}$ for at least one of the two EEG-channels. Let $r = 1, \dots, 52$ be the index for the rolling window. The rolling window of 2s can be moved on 52 times until the maximum duration of 15s is reached. Let $l = 0.25s$ be the length of each step and k^* be a detected start second. In each step the following condition is tested:

$$\frac{\sum_{b=\theta,\alpha,\beta} P_b(I_{+2}(k^* + (l \cdot r)))}{\sum_{b=\theta,\alpha,\beta} P_b(I_{-10}(k^*))} > threshold_{duration}$$

Arousal duration is added in 0.25s steps as long as this condition is fulfilled for at least one of the two EEG-channels and maximum duration of 15s is not reached yet.

2.2.5.3 Verification of further arousal criteria

Check for an arousal duration of at least 3 seconds

The definition of the AASM states that an arousal must last at least 3s. Hence, all detected arousals that don't meet this criterion are excluded. Doing so, some EEG amplitude artifacts, that provoke very high values (outliers) with a high frequency for a short time, are excluded.

Check for 10s of stable sleep before an arousal start detection

After removing all arousals shorter than 3s, a routine is implemented to verify if an arousal is not followed by another arousal within the following 10s after the end of the first. All detections within the first 10s after the end of the last detection are removed.

Check for arousals in REM-sleep

To verify the correct detection of arousals during REM-sleep a check for an increase in submental EMG (chin EMG) that lasts at least 1s is necessary. For all detections that appear to take place while the hypnogram is equal to 5 (which indicates REM sleep) the chin EMG is tested. There is no definition on what is exactly meant by an increase in chin EMG neither there is a note on when exactly the increase in chin EMG must appear. In practice the increase has to be around the start of an arousal, which means approximately not earlier than 1s before the start and 2s after the start. The second open question here is how to measure an increase in submental EMG. Here, a simple amplitude check within the interval of 1s before the start until 2s after the detected start, is implemented.

The routine is checking if the chin EMG within a 3s interval $I_{+3}(k^* - 1)$ starting 1s before the start second of a detected arousal k^* and ending 2s after k^* is greater than 2x standard deviation of the whole recording of chin EMG (with length T):

$$\max_{t \in I_{+3}(k^*-1)} \left(chin(t) - \overline{chin}\right) > 2 \cdot \sqrt{\sum_{t=1}^{T} \left(chin(t) - \overline{chin}\right)^2}$$

2.2.6 Validation of detection algorithm

The validation of the whole detection algorithm differs from the validation of the classifier. Hence, the already explained statistical terms have to be redefined for this validation problem:

- TP (True Positives) = Number of annotated arousals that are overlapped by at least one detection
- FP (False Positives) = Number of detected arousals that are not overlapping with any annotated arousal
- FN (False Negatives) = Number of annotated arousals that are not not overlapping with any detection
- TN (True Negatives) = There is no useful measure for TN when considering whole arousal as they could last from 3s up to 15s or in manual annotations sometimes even up to 30s.

To validate the results, it is tested if the detected arousals after defining their start and their end are overlapping with any annotated arousal. Since it's nearly impossible to detect the exact position of the expert annotated arousal, an arousal is seen as correctly detected if any overlapping period is detected (True Positives). False Negatives are all not detected, but annotated arousals.

The following statistical values are considered for validation:

- *True Positive Rate (Sensitivity):* $\frac{TP}{TP+FN}$

(Percentage of all true arousals that were correctly detected - capability of the process to detect arousals*)*

- *False Negative Rate:* $\frac{FN}{TP+FN} = 1 - sensitivity$
  (Percentage of all true arousals that were not correctly detected)
- *Positive predictive value:* $\frac{TP}{TP+FP}$
  (Percentage of all detected arousals that are true arousals)

All statistical values mentioned in the section *validation of classifier* that need True Negatives to be calculated, are left out in this validation.

The results are obtained by applying Leave-One-Out-Cross validation. For each patient, the SVM is trained with 6 patients and tested on the remaining one patient.

2.3 Methods for the investigation on the relation between arousals and leg movements

The aim of this section is to investigate the relation between arousals and leg movements.

From a medical point of view the occurrence of leg movements is more important the more it leads to disturbance of sleep and problems with frequent awakenings, which can lead to daytime sleepiness as well as further problems. Arousals can be an indicator of the quality of sleep of the patient. Thus, it is of high interest to investigate the relation between both events, to be able to make a statement about how frequent awakenings are related to leg movements.

The complex relation between these two events has already been researched, for example in the work of (Ferri, et al., 2015) or the work of (Hornyak, Feige, Voderholzer, & Riemann, 2005). The outcome of these studies is that there might be a more complex mechanism behind the relationship of both events rather than a simple causality.

For better understanding, the definition of the AASM for the association between a PLM and an arousal is repeated here:

An arousal and a limb movement that occur in a PLM series should be considered associated with each other if they occur simultaneously, overlap, or when there is <0.5 seconds between the end of one event and the onset of the other event regardless of which is first. (Berry, et al., 2016)

2.3.1 Data set

The association of leg movements and arousals can only be tested reasonably if there are enough leg movements as well as arousals in the recording. Hence, for this research the tested data is a subset of the 7 patients of AKH Wien, consisting only of the 4 patients with IDs TS100714, TS090714, TC070814, MB200814, who are diagnosed with PMLS. To get a better understanding of the data set, several indices and sleep architecture parameters are calculated for each patient in section *Results*.

The annotated leg movements in the PSG data of these patients are already classified as PLM and LM. For better understanding it will always be referred to leg movements (LM) as the sum of periodic limb movements (PLM) and isolated limb movements (iLM). In this work the above repeated definition of the association between a PLM and an arousal will also be applied to iLMs.

The marked area in Figure 33 shows an annotated LM of both legs (iLM) that is associated with an arousal.

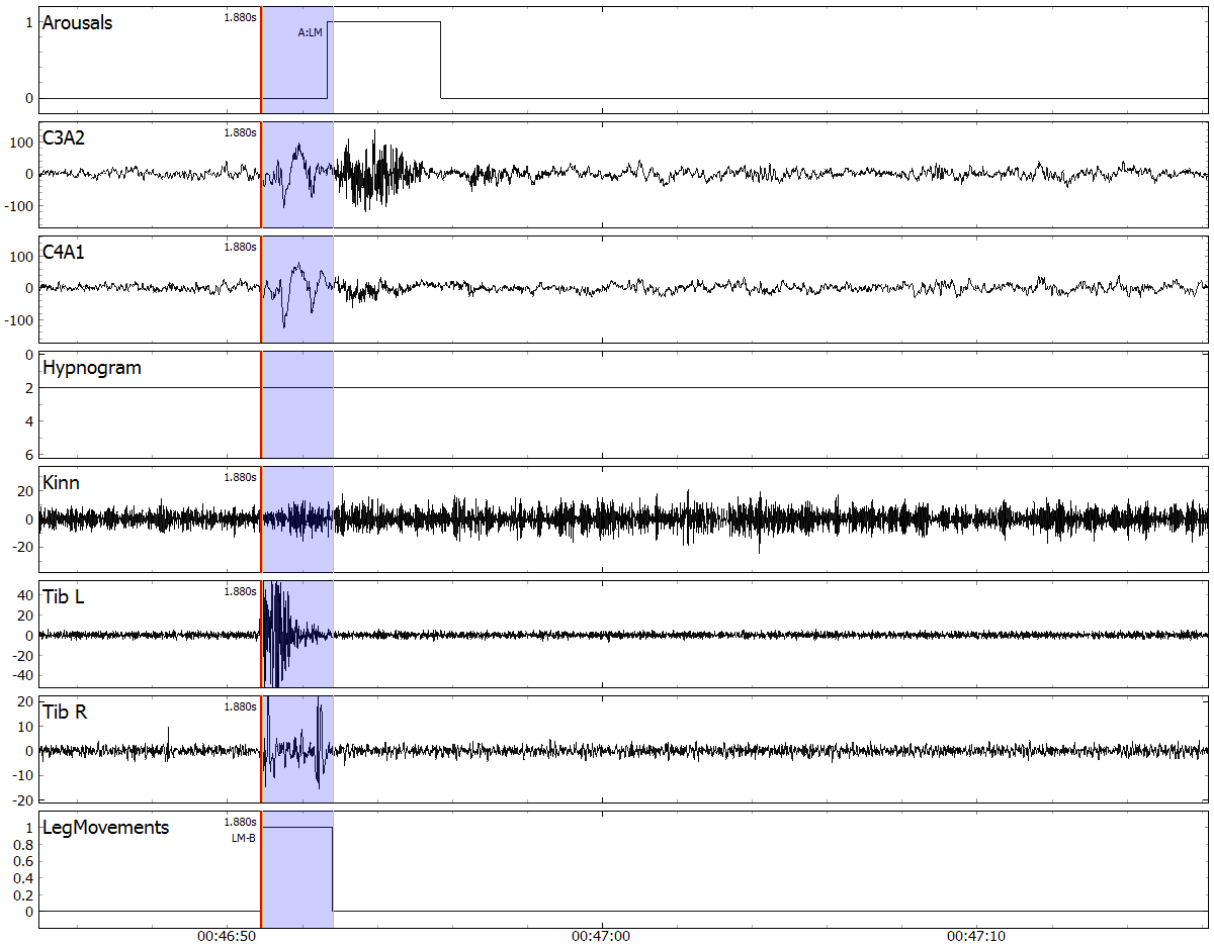


Figure 33: Illustration of a LM associated with an arousal, visualized with the "Sleep-Data-Viewer-Tool" of the AIT; 1. Line: Arousal annotations; 2. Line: EEG channel C3A2; 3. Line: EEG channel C4A1; 4. Line: Hypnogram; 5. Line: Chin EMG; 6. Line: EMG left leg; 7. line: EMG right leg; 8. Line: Leg Movement annotations

Furthermore, the intensity of a leg movement will be measured with two different approaches: On the one hand with the annotated duration in the PSG data and on the other hand with a value computed by the 3D detection algorithm of the *AIT.*

A leg movement as well as an arousal is described by its start time and its duration. For further analysis, the start times will be notated with capital letters and the duration with small letters:

- L for start time and l for duration of leg movements
- A for start time and a for duration of arousals

Let $L = \{L_1, \dots, L_N\}$ be the set of all leg movements start times within all recordings of the 4 patients, indexed with the set $I = \{1, \dots, N\}$. Each start time L_i is associated with the duration l_i, $i \in I$. Let I_{PLM} be a subset of I describing all PLMs and let I_{iLM} be a subset of I describing all iLMs, so that $I = \{1, \dots, N\} = I_{PLM} \cup I_{iLM}$. The start times of PLMs and iLMs can be described as subsets of the start times of all LMs: $PLM\ start\ times = \{L_i \mid i \in I_{PLM}\}$ and $iLM\ start\ times = \{L_i \mid i \in I_{iLM}\}$. The cardinality of the set X is described by $|X|$, for example: $|I| = N$.

For the arousal annotations, both sets of annotations, the original and the modified one are considered. In the following, PLMs as well as iLMs are considered to be part of an (P/i)LM-Arousal pair, when the criterion of the above repeated definition of the association of the two events is met.

Let $I_{LM_{pair}}$, $I_{PLM_{pair}}$ and $I_{iLM_{pair}}$ also be subsets of I to describe which LM/PLM/iLM is associated with an arousal.

Let arousal start times $A = \{A_1, \dots, A_K\}$ be indexed with $J = \{1, \dots, K\}$, with K being the number of arousals in all considered recordings. Each start time A_j is associated with the duration $a_j, j = 1, \dots, K$. Furthermore let $j(i)$ be the index to describe the associated arousal time or duration for leg movements indexed with $i \in I$.

2.3.2 Analysis of occurrence of associated LM and arousal events

To get an idea of what part of arousals and LMs is associated to each other, various parameters are calculated for both arousal annotation sets.

- Total number of LMs/ arousals:
  - LMs: $|I|$, PLMs: $|I_{PLM}|$, iLMs: $|I_{iLM}|$
  - Arousals: $|J|$
- Part of PLMs/iLMs out of all LMs:
  - PLM: $\frac{|I_{PLM}|}{|I|}$, iLM: $\frac{|I_{iLM}|}{|I|}$
- Total number of LM-Arousal pairs:
  - LMs: $\left|I_{LM_{pair}}\right|$, PLMs: $\left|I_{PLM_{pair}}\right|$, iLMs: $\left|I_{iLM_{pair}}\right|$,
- Part of LM-Arousal pairs out of all LMs/Arousals:
  - LMs: $\frac{\left|I_{LM_{pair}}\right|}{|I|}, \frac{\left|I_{LM_{pair}}\right|}{|J|}$
  - PLMs: $\frac{\left|I_{PLM_{pair}}\right|}{|I|}, \frac{\left|I_{PLM_{pair}}\right|}{|J|}$
  - iLMs: $\frac{\left|I_{iLM_{pair}}\right|}{|I|}, \frac{\left|I_{iLM_{pair}}\right|}{|J|}$

2.3.3 Analysis of cause/effect relation between LM and arousal events

Next to the question which part of all events is associated with each other, another interesting approach is to analyze a possible causality between them. Meaning which event causes the other in terms of earlier start time. For this analysis, only associated LM-arousal pairs are considered. Hence, the difference in start times of associated events is calculated:

$$L_i - A_{j(i)} > 0 \; if \; arousal \; precedes \; LM$$

$$L_i - A_{j(i)} < 0 \; if \; LM \; precedes \; arousal$$

In the chapter *Results* the distribution of the difference in start times of associated events is plotted with a histogram and it is tested for normal distribution. In addition, a statistical test is performed to test if the mean value of the difference in start times is significantly different from zero which would indicate that a certain event starts statistically significant more often before the other event.

2.3.4 Analysis of relation in duration between LM and arousal events

As already mentioned in the work of (Ferri, et al., 2015) the relation between the duration of associated LM and arousal events can be investigated. A graphical analysis with scatterplots and a regression line has been calculated by estimating the parameters C and β in the model:

$$a_{j(i)} = C + \beta \cdot l_i + u_t$$

2.3.5 Analysis of relation between the intensity of a LM and the occurrence of an arousal

Another hypothesis is that more intense leg movements are more likely to cause arousals than others. In this context, it has to be defined how the intensity of a LM can be measured. Two main ideas are realized in this work: On the one hand the duration of a LM in the PSG data and on the other hand a value that is computed with 3D detection of LMs in the *Austrian Institute of Technology* are tested.

In the following the term "3D detection" will always refer to the 3D detection of LMs.

At first the duration is considered as a measure for the intensity: The Operator \ means "without". For example, $I \backslash I_{LM_{pair}}$ is the index set of all LMs without the indices for LMs associated with an arousal.

The mean value of the duration of LM/arousals (not) associated with arousals is calculated:

- Mean of duration of leg movements not associated with an arousal: $\frac{1}{|I \backslash I_{LM_{pair}}|} \sum_{i \in I \backslash I_{LM_{pair}}} l_i$
- Mean of duration of leg movements associated with an arousal: $\frac{1}{|I_{LM_{pair}}|} \sum_{i \in I_{LM_{pair}}} l_i$
- Mean of duration of PLMs not associated with an arousal: $\frac{1}{|I_{PLM} \backslash I_{PLM_{pair}}|} \sum_{i \in I_{PLM} \backslash I_{PLM_{pair}}} l_i$
- Mean of duration of PLMs associated with an arousal: $\frac{1}{|I_{PLM_{pair}}|} \sum_{i \in I_{PLM_{pair}}} l_i$
- Mean of duration of iLMs not associated with an arousal: $\frac{1}{|I_{iLM} \backslash I_{iLM_{pair}}|} \sum_{i \in I_{iLM} \backslash I_{iLM_{pair}}} l_i$
- Mean of duration of iLMs associated with an arousal: $\frac{1}{|I_{iLM_{pair}}|} \sum_{i \in I_{iLM_{pair}}} l_i$

One can say that the duration of different leg movements is independent from each other and therefore statistical tests are valid to check if there is a significant difference between the mean within the leg movement groups associated with arousal and not associated with arousal. A two-sided t-test on a significant difference in the mean within the two groups is performed and a graphical analysis with boxplots is provided in the chapter results.

After investigating a possible relation between the duration of an event and its association with an arousal, the so called "area" value of the 3D detection is considered as a measure for the intensity of a leg movement. In 3D detection, the pixels in the ROI (Region of interest (2D picture): legs) are observed with a certain sample rate and the standard deviation of the intensity of each pixel in time, considering the 15 frames before and the 15 frames after a certain point in time, is computed. As long as this standard deviation remains over a certain threshold an event is detected. The "area" value is calculated by summing up all standard deviation values during a 3D detection. It can be interpreted as a possible value to describe how "big" or how "strong" a LM is.

The base of all investigations in this work is the PSG data of certain patients of AKH Wien. It is possible that the 3D detection algorithms detect more than one event during one annotation in the PSG data, or that sometimes there are PSG annotations that aren't detected at all by the 3D detection. Area values are assigned to LMs according to the following rules:

- If a LM is overlapped by one or more 3D detections, the sum of all area values of 3D detections that overlap the annotation is considered as the area value for the LM.
- If a LM is not detected by 3D detection, the area value is set to zero for this LM.

The 3D detection is validated for the considered patients. Let the exponent "$3D$" indicate if a LM indexed with $I = \{1, \dots, N\}$ is detected by the 3D detection. The index I^{3D} describes the LMs that are detected by 3D detection. The same definition can be applied to all already defined subsets of I.

The following numbers are calculated to analyze the abilities of the 3D detection on the used data set:

- Total number of 3D detected and not 3D detected LMs:
  - $|I^{3D}|, |I \backslash I^{3D}|$
- Percentage of 3D detected and not 3D detected LMs out of all LMs:
  - $\frac{|I^{3D}|}{|I|}, \frac{|I \backslash I^{3D}|}{|I|}$
- Total number of all LMs/PLMs/iLMs with (in pair) and without arousals that are 3D detected and not 3D detected:
  - 3D detected with arousals: LMs: $\left|I^{3D}_{LM_{pair}}\right|$, PLMs: $\left|I^{3D}_{PLM_{pair}}\right|$, iLMs: $\left|I^{3D}_{iLM_{pair}}\right|$
  - 3D detected without arousals: LMs: $\left|I^{3D} \backslash I^{3D}_{LM_{pair}}\right|$, PLMs: $\left|I^{3D}_{PLM} \backslash I^{3D}_{PLM_{pair}}\right|$, iLMs: $\left|I^{3D}_{iLM} \backslash I^{3D}_{iLM_{pair}}\right|$
  - Not 3D detected with arousals: LMs: $\left|I_{LM_{pair}} \backslash I^{3D}_{LM_{pair}}\right|$, PLMs: $\left|I_{PLM_{pair}} \backslash I^{3D}_{PLM_{pair}}\right|$, iLMs: $\left|I_{iLM_{pair}} \backslash I^{3D}_{iLM_{pair}}\right|$
  - Not 3D detected without arousals: LMs: $\left|I \backslash (I^{3D} \cup I_{LM_{pair}})\right|$, PLMs: $\left|I_{PLM} \backslash (I^{3D}_{PLM} \cup I_{PLM_{pair}})\right|$, iLMs: $\left|I_{iLM} \backslash (I^{3D}_{iLM} \cup I_{iLM_{pair}})\right|$
- Percentage of all LMs/PLMs/iLMs with (in pair) and without arousals that are 3D detected and not 3D detected:
  - % of 3D detected with arousals: LMs: $\frac{\left|I^{3D}_{LM_{pair}}\right|}{|I^{3D}|}$, PLMs: $\frac{\left|I^{3D}_{PLM_{pair}}\right|}{|I^{3D}_{PLM}|}$, iLMs: $\frac{\left|I^{3D}_{iLM_{pair}}\right|}{|I^{3D}_{iLM}|}$
  - % of 3D detected without arousals: LMs: $\frac{\left|I^{3D} \backslash I^{3D}_{LM_{pair}}\right|}{|I^{3D}|}$, PLMs: $\frac{\left|I^{3D}_{PLM} \backslash I^{3D}_{PLM_{pair}}\right|}{|I^{3D}_{PLM}|}$, iLMs: $\frac{\left|I^{3D}_{iLM} \backslash I^{3D}_{iLM_{pair}}\right|}{|I^{3D}_{iLM}|}$
  - % of not 3D detected with arousals: LMs: $\frac{\left|I_{LM_{pair}} \backslash I^{3D}_{LM_{pair}}\right|}{|I \backslash I^{3D}|}$, PLMs: $\frac{\left|I_{PLM_{pair}} \backslash I^{3D}_{PLM_{pair}}\right|}{|I_{PLM} \backslash I^{3D}_{PLM}|}$, iLMs: $\frac{\left|I_{iLM_{pair}} \backslash I^{3D}_{iLM_{pair}}\right|}{|I_{iLM} \backslash I^{3D}_{iLM}|}$

- % of not 3D detected without arousals: LMs: $\frac{\left|I \backslash (I^{3D} \cup I_{LM_{pair}})\right|}{|I \backslash I^{3D}|}$, PLMs: $\frac{\left|I_{PLM} \backslash (I_{PLM}^{3D} \cup I_{PLM_{pair}})\right|}{|I_{PLM} \backslash I_{PLM}^{3D}|}$, iLMs: $\frac{\left|I_{iLM} \backslash (I_{iLM}^{3D} \cup I_{iLM_{pair}})\right|}{|I_{iLM} \backslash I_{iLM}^{3D}|}$

It will be tested graphically and statistically if the mean and the median of both intensity values are significantly different within LMs that are associated with arousals and LMs that aren't associated with arousals. For the intensity measured in area values, additionally an analysis on the intensity of LMs with and without arousals categorized in 4 intensity classes is done and visualized. The categories are defined by calculating the 25%, 50% and the 75% quantile of the distribution of all area values for all 3D detected LMs and assigning a category to the area values according to the following rule:

- Category = 0 if area == 0 (not 3D detected LM)
- Category = 1 if area $\in (0, q_{0.25}^{3D})$
- Category = 2 if area $\in [q_{0.25}^{3D}, q_{0.5}^{3D})$
- Category = 3 if area $\in [q_{0.5}^{3D}, q_{0.75}^{3D})$
- Category = 4 if area $\in [q_{0.75}^{3D}, \max(area)]$

To see in which category most of the LMs with or without arousals are placed, histograms of the categorized "area" values are plotted.

3 Results

This chapter will be concerned with the presentation of the results of the arousal detection algorithm and the investigation on the relation between arousals and leg movements.

3.1 Results of the arousal detection

All steps of the arousal detection are performed with the modified annotation set.

3.1.1 Validation of Classifier

As described in Section *Validation of Classifier* the Parameters of the SVM are chosen by performing a grid search and determining the best parameters by maximizing the Youden-Index.

The following parameters were chosen: $C = 11, \gamma = 2^{-12}$

Features: $e_\theta, e_\alpha, e_\beta; d_{\theta vs(\alpha+\beta)}; \tau_\theta, \tau_\alpha, \tau_\beta; ar(6): coefficients\ for\ lag1 - lag6$ (each for both channels, resulting in a total of 26 features)

The validation is done by applying Leave-One-Out-Cross validation, which means that for each patient results are received by training the SVM on 6 patients and testing it on the one remaining patient. Table 3 displays the results of the classifier for the detection of arousal start segments per patient. The last column (in grey) of Table 4 is counting the annotated arousals that are within the 30s before an epoch that is staged as awakening. As described in Check of hypnogram these periods are excluded because of the confusion with “transition to wake” events. Nevertheless, some arousals are scored in these epochs and for completeness they are mentioned here. When calculating statistical values they are not included, because they can’t be detected due to the exclusion criteria. In further tables the number of arousals (start segments) annotated in the 30s before awakening are notated in brackets and in grey next to the total number of annotated arousals (start segments), for instance see Table 3.

Modified annotations:

	Number of annotated start segments	Total number of segments	Number of detected start segments	Sensitivity	False Negative Rate	Positive Predictive value	Specificity	Accuracy	False Positive Rate	Negative Predictive Value
MB200814	83 (1)	8505	294	96%	4%	27%	97%	97%	3%	100%
CP101214	11 (0)	9350	41	91%	9%	24%	100%	100%	0%	100%
MC250614	68 (1)	9270	406	99%	1%	16%	96%	96%	4%	100%
TS100714	205 (2)	6770	722	98%	2%	28%	92%	92%	8%	100%
MP020714	53 (2)	7550	186	96%	4%	26%	98%	98%	2%	100%
TS090714	201 (4)	8060	733	98%	2%	26%	93%	93%	7%	100%
TC070814	79 (1)	9650	322	96%	4%	23%	97%	97%	3%	100%
Sum/Mean	***700 (11)***	***59155***	***2704***	***96%***	***4%***	***24%***	***96%***	***96%***	***4%***	***100%***

Table 3: Results of validation of arousal start detection (modified annotations)

In addition, an analysis of the overall detection performance of arousal events is done. Therefore, the statistical values are calculated on the total TP/FN/FP/TN instead of per patient.

	True Negatives	True Positives	False Positives	False Negatives	Arousals in 30s before awakening
MB200814	8208	79	215	3	1
CP101214	9308	10	31	1	0
MC250614	8863	66	340	1	1
TS100714	6044	199	523	4	2
MP020714	7362	49	137	2	2
TS090714	7324	194	539	3	4
TC070814	9325	75	247	3	1
Sum	***56434***	***672***	***2032***	***17***	***11***

Table 4: Validation results of overall start detection performance

Sensitivity	False Negative Rate	Positive Predictive Value	Specificity	Accuracy	False Positive Rate	Negative Predictive Value
97,53%	2,47%	24,85%	96,52%	96,54%	3,48%	99,97%

Table 5: Statistical values for overall arousal start segment (3s) detection performance

3.1.2 Validation of arousal start area

For the final determination of arousals, the detected start segments that are located directly next to each other are joined and further analyzed as described in *Final arousal determination.* The following table shows how many start areas are detected, that are further analyzed for final determination of definite arousal. When considering variable length start areas the calculation of True Negatives doesn't make any sense and for that reason only statistical values without necessity of True Negatives are calculated. One has to note that the Sensitivity and the False Negative Rate don't change when joining start segments to start areas. Only the Positive Predictive Value is increasing because False Positive segments that are joined with a True Positive segment are no longer counted as a False Positive segment, but part of one start area that is overlapping with a true start segment.

	Number of annotated arousals	Number of detected start areas	Sensitivity	False Negative Rate	Positive Predictive Value
MB200814	83 (1)	187	96%	4%	42%
CP101214	11 (0)	21	91%	9%	48%
MC250614	68 (1)	250	99%	1%	26%
TS100714	205 (2)	368	98%	2%	54%
MP020714	53 (2)	105	96%	4%	47%
TS090714	201 (4)	397	98%	2%	49%
TC070814	79 (1)	181	96%	4%	41%
Sum/Mean	***700 (11)***	***1509***	***96%***	***4%***	***44%***

Table 6: Validation results after joining detected start segments next to each other

3.1.3 Validation of final arousals

After classifying arousal starts with the SVM and joining segments to start areas, the second step of the detection algorithm is done. The threshold for the calculation of the duration of an arousal is set at: $threshold_{duration} = 2$. After computing the duration of the possible arousal, all arousals that

occur within the 10s after the last arousal are removed. Furthermore, all detections that happen to be shorter than 3s are also removed and last the check for the change in submental EMG, while REM sleep arousals, is done.

As described in Section *Validation of detection algorithm* all values that make use of True Negatives can't be calculated here. The remaining statistics are displayed in Table 7.

Modified annotations:

	Number of annotated arousals	Number of detected arousals	Sensitivity	False Negative Rate	Positive Predictive Value
MB200814	83 (1)	116	87%	13%	61%
CP101214	11 (0)	15	91%	9%	67%
MC250614	68 (1)	171	88%	12%	35%
TS100714	205 (2)	239	84%	16%	71%
MP020714	53 (2)	68	78%	22%	59%
TS090714	201 (4)	228	82%	18%	71%
TC070814	79 (1)	136	92%	8%	53%
Sum/Mean	***700 (11)***	***973***	***86%***	***14%***	***60%***

Table 7: Results of validation of final arousals (modified annotations)

	True Positives	False Positives	False Negatives	Arousals in 30s before awakening
MB200814	71	45	11	1
CP101214	10	5	1	0
MC250614	59	112	8	1
TS100714	170	69	33	2
MP020714	40	28	11	2
TS090714	162	66	35	4
TC070814	72	64	6	1
Sum	***584***	***389***	***105***	***11***

Table 8: Validation results of overall final arousal detection performance

Sensitivity	Positive Predictive Value	False Negative Rate
84,76%	60,02%	15,24%

Table 9: Statistical values for overall final arousal detection performance

For the final determination of definite arousals, various steps were performed according to the definition of the AASM criteria. It can be noted that sensitivity decreased on average by about 10%. This is because neither the end of an arousal nor the term "stable sleep" are defined clearly by the AASM and so the decision for final arousals is again rather subjective. The implemented algorithms are configured to perform best in terms of high sensitivity and high positive predictive value. To improve the performance of the overall detection algorithms, especially the second step needs to be analyzed regarding a better understanding of the two critical points within the definition: the end of an arousal and the exact definition of stable sleep.

3.1.4 Detection examples

In the following, plots for detections and their features are provided for patient TS100714. The red area labeled with expert arousal is always referring to modified annotations as the algorithm is trained, tested and validated on this set of annotations. More plots of all kinds can be found in the Appendix

True Positives

Filtered Signal:

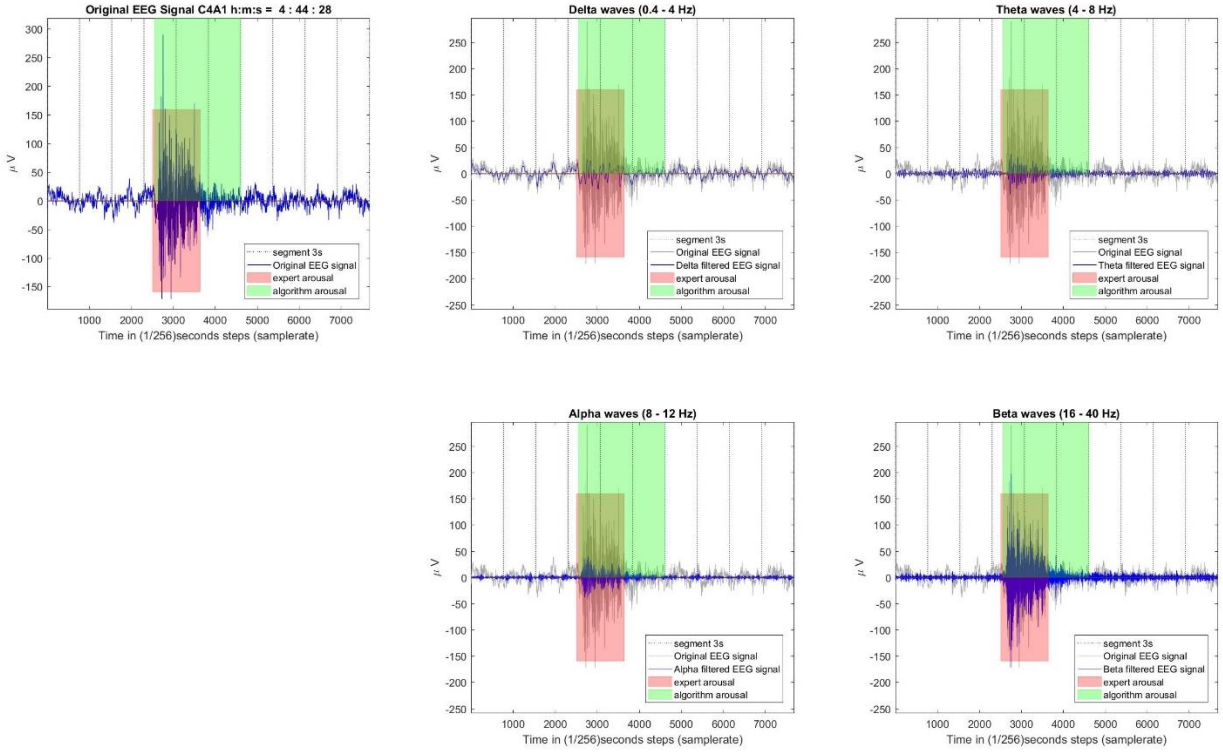


Figure 34: Illustration of the filtered signal during a True Positive detection; EEG-channel C4A1

Features:

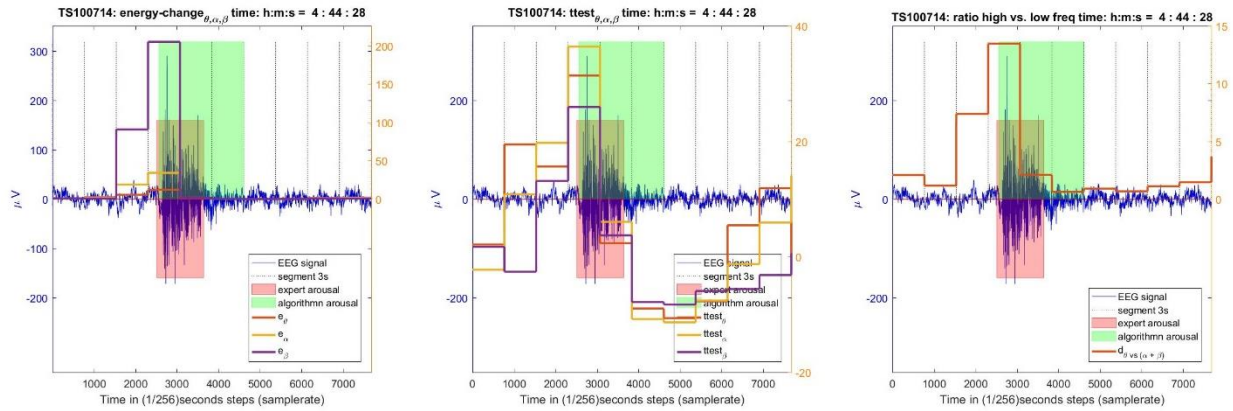


Figure 35: Illustration of the behavior of different feature sets, used for classification of arousal start segments, during a True Positive detection

AR-Features:

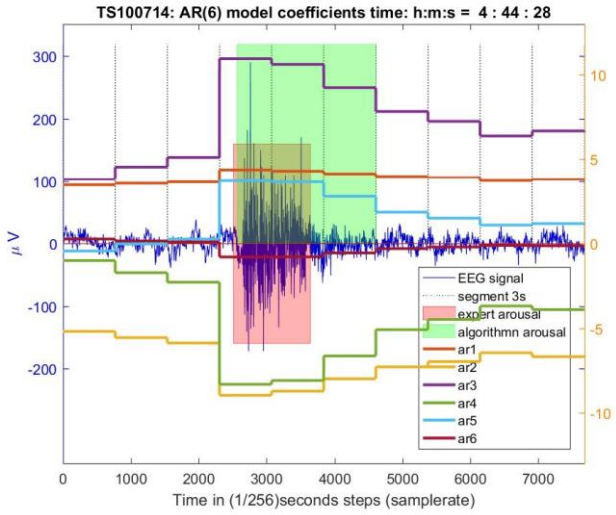


Figure 36: Illustration of the behavior of the feature set of coefficients of an AR(6) model, used for classification of arousal start segments, during a True Positive detection

False Negatives

Filtered signal:

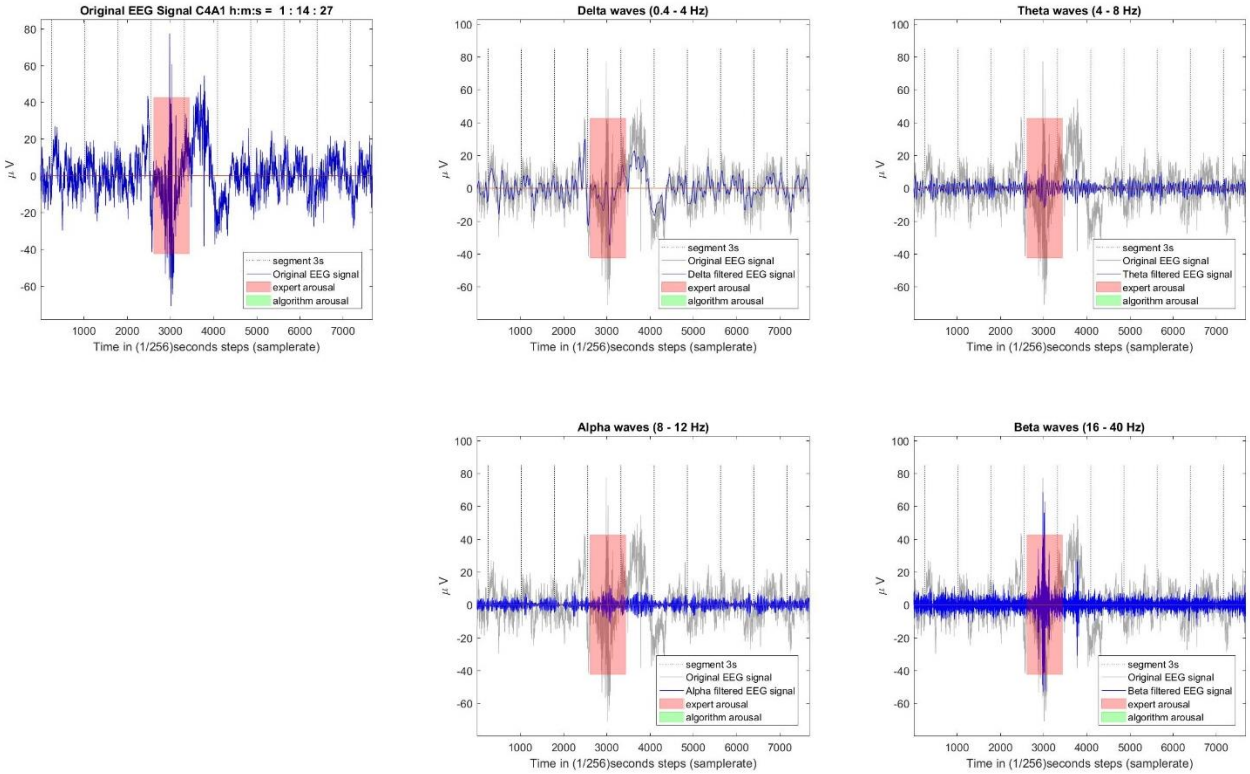


Figure 37: Illustration of the filtered signal during a False Negative detection; EEG-channel C4A1

Features:

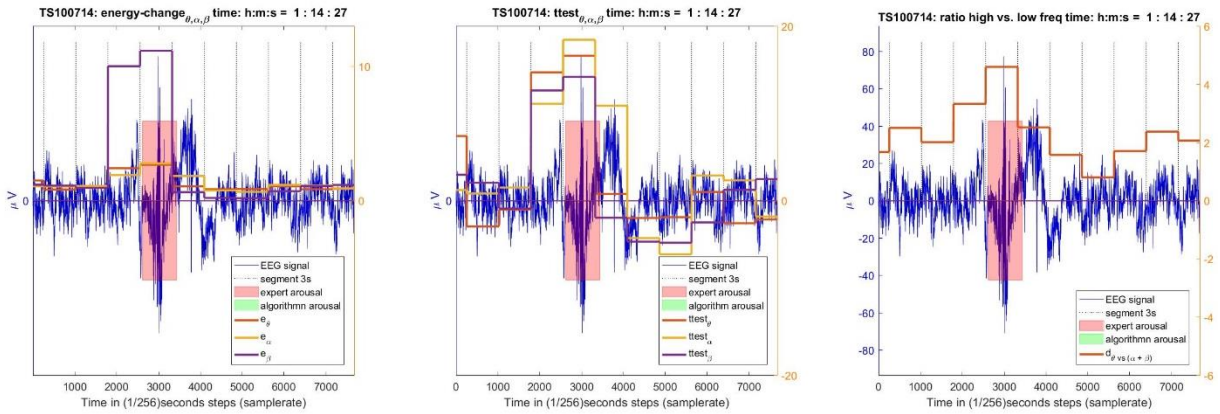


Figure 38: Illustration of the behavior of different feature sets, used for classification of arousal start segments, during a False Negative detection

AR-Features:

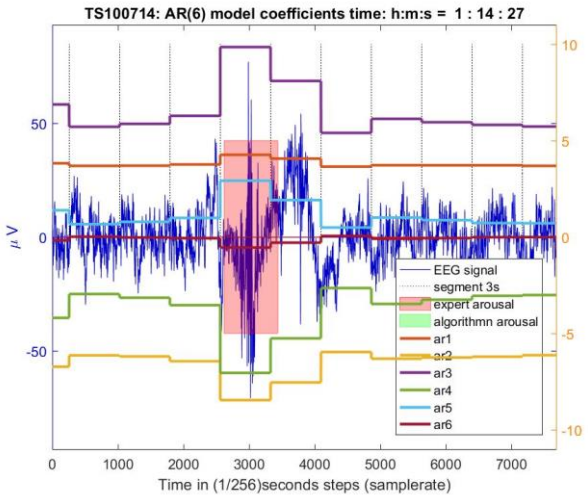


Figure 39: Illustration of the behavior of the feature set of coefficients of an AR(6) model, used for classification of arousal start segments, during a False Negative detection

False Positives

Filtered signal:

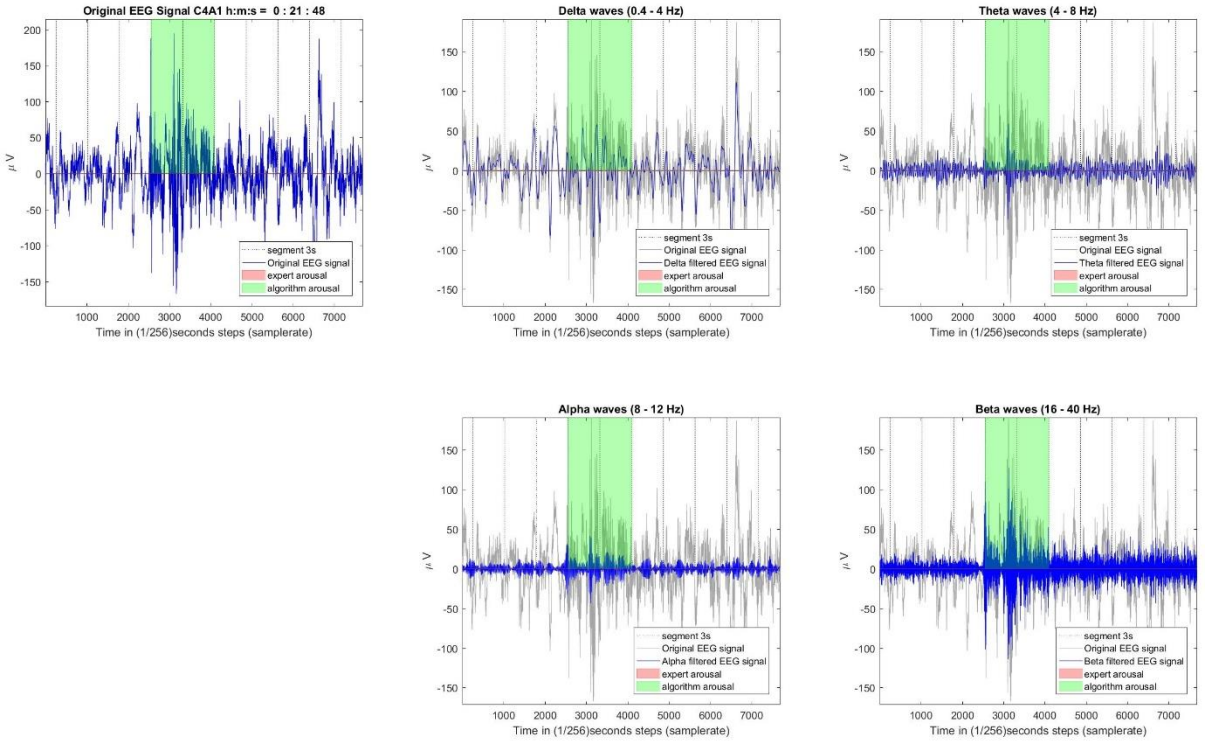


Figure 40: Illustration of the filtered signal during a False Positive detection; EEG-channel C4A1

Features:

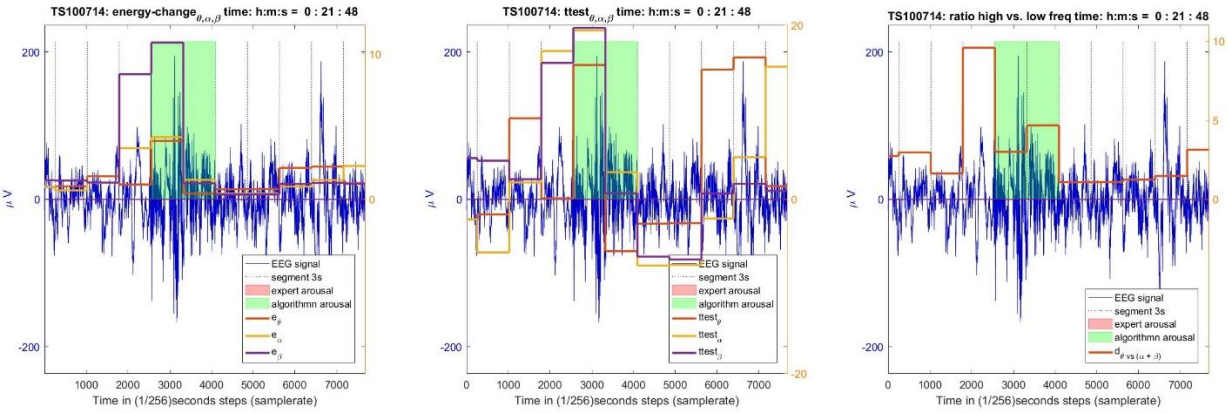


Figure 41: Illustration of the behavior of different feature sets, used for classification of arousal start segments, during a False Positive detection

AR-Features:

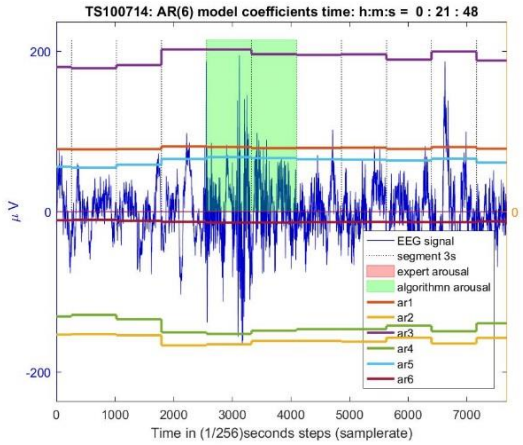


Figure 42: Illustration of the behavior of the feature set of coefficients of an AR(6) model, used for classification of arousal start segments, during a False Positive detection

3.2 Results of the investigation on the relation between arousals and leg movements

This section is concerned with the presentation and visualization of the results of the investigations on the relation between leg movements and arousal events. Firstly, a short summary of different indices and sleep architecture parameters is given for the set of considered patients with PLMS. Here all annotated arousals are included, also the ones occurring in the 30s before an awakening. In the following, the term "TST" describes the *Total Sleep Time*, a measure for the time in which the patient was sleeping during the recording.

Original annotations:

	TS100714	TS090714	TC070814	MB200814
TST in h	6,2	7,2	8,3	7,3
Total Limb Movement Index (per h of TST):	65,9	70,3	14,2	37,0
Periodic Limb Movement Index (per h of TST):	**55,2**	**56,5**	**4,5**	**33,5**
Isolated Limb Movement Index (per h of TST):	10,7	13,8	9,8	3,4
Arousal Index (per h of TST):	38,4	32,7	13,6	15,8
Periodic Limp Movements paired with an Arousal Index (per h of TST):	17,9	14,5	1,7	6,6
Total number of Arousals (in recording):	238	235	113	115
Total number of Limb Movements (in recording):	408	505	118	269
Total number of Isolated Limb Movements (in recording):	66	99	81	25
Total number of Periodic Limb Movements (in recording):	342	406	37	244
Total number of LM-Arousal pairs (in recording):	142	149	47	62
Total number of PLM-Arousal pairs (in recording):	118	111	14	49
Total number of iLM-Arousal pairs (in recording):	24	38	33	13
Part of PLMs out of all LMs:	84%	80%	31%	91%
Part of isolated LMs out of all LMs:	16%	20%	69%	9%

Table 10: Indices and sleep architecture parameters for considered patients with PLMS diagnosis; original arousal annotations

As it can be seen in Table 10 the PLMI (Periodic Limb Movement Index) is close to 5 or greater than 5 for all patients, which indicates a PLMS diagnosis. It can be observed, that patient TC070814 has an PLMI of only 4,5 (see Table 10) and shows rather different sleep architecture parameters than the other 3 patients. Nevertheless, the patient is included for further analysis because of the fact that there are still 47 LM-Arousal pairs, which can be used for investigation.

Modified annotations:

	TS100714	TS090714	TC070814	MB200814
TST in h	6,2	7,2	8,3	7,3
Total Limb Movement Index (per h of TST):	65,9	70,3	14,2	37,0
Periodic Limb Movement Index (per h of TST):	**55,2**	**56,5**	**4,5**	**33,5**
Isolated Limb Movement Index (per h of TST):	10,7	13,8	9,8	3,4
Arousal Index (per h of TST):	33,1	28,0	9,5	11,4
Periodic Limp Movements paired with an Arousal Index (per h of TST):	16,5	12,4	1,2	5,8
Total number of Arousals (in recording):	205	201	79	83
Total number of Limb Movements (in recording):	408	505	118	269
Total number of Isolated Limb Movements (in recording):	66	99	81	25
Total number of Periodic Limb Movements (in recording):	342	406	37	244
Total number of LM-Arousal pairs (in recording):	120	119	31	53
Total number of PLM-Arousal pairs (in recording):	102	89	10	42
Total number of iLM-Arousal pairs (in recording):	18	30	21	11
Part of PLMs out of all LMs:	84%	80%	31%	91%
Part of isolated LMs out of all LMs:	16%	20%	69%	9%

Table 11: Indices and sleep architecture parameters for considered patients with PLMS diagnosis; modified arousal annotations

3.2.1 Analysis of occurrence of associated LM and arousal events

As described in section *Methods for the investigation on the relation between arousals and leg movements*, the relevant values for the occurrence of associated arousal events are calculated. There is a differentiation between the original annotations and the modified annotations for arousals.

Original annotations:

	TS100714	TS090714	TC070814	MB200814
Part of LM-Arousal pairs out of all LMs:	**35%**	**30%**	**40%**	**23%**
Part of PLM-Arousal pairs out of all PLMs:	35%	27%	38%	20%
Part of iLM-Arousal pairs out of all iLMs:	36%	38%	41%	52%
Part of LM-Arousal pairs out of all Arousals:	**60%**	**63%**	**42%**	**54%**
Part of PLM-Arousal pairs out of all Arousals:	50%	47%	12%	43%
Part of iLM-Arousal pairs out of all Arousals:	10%	16%	29%	11%

Table 12: Statistics for the occurrence of associated events with original arousal annotations; Approximation for the conditional probability of the occurrence of an arousal given that an LM appears and vice versa.

Modified annotations:

	TS100714	TS090714	TC070814	MB200814
Part of LM-Arousal pairs out of all LMs:	**29%**	**24%**	**26%**	**20%**
Part of PLM-Arousal pairs out of all PLMs:	30%	22%	27%	17%
Part of iLM-Arousal pairs out of all iLMs:	27%	30%	26%	44%
Part of LM-Arousal pairs out of all Arousals:	**59%**	**59%**	**39%**	**64%**
Part of PLM-Arousal pairs out of all Arousals:	50%	44%	13%	51%
Part of iLM-Arousal pairs out of all Arousals:	9%	15%	27%	13%

Table 13: Statistics for the occurrence of associated events with modified arousal annotations; Approximation for the conditional probability of the occurrence of an arousal given that an LM appears and vice versa.

In Table 12 and Table 13 it can be observed, that in three patients with both arousal annotations sets more than 50% out of all arousals are associated with any LM. This is a first indicator for the existence of some relation between these events.

By summarizing all statistics for the group of the 4 PLMS patients, the following values are obtained (all annotated arousals are included, also the ones occurring in the 30s before an awakening):

	Original annotations	Modified annotations
Total number of Arousals (in all recordings):	701	568
Total number of LMs (in all recordings):	1300	1300
Total number of iLMs (in all recordings):	271	271
Total number of PLMs (in all recordings):	1029	1029
Part of PLM out of all LMs	79%	79%
Part of iLMs out of all LMs	21%	21%
Total number of LM-Arousal pairs (in all recordings):	400	323
Total number of PLM-Arousal pairs (in all recordings):	292	243
Total number of iLM-Arousal pairs (in all recordings):	108	80
Part of LM-Arousal pairs out of all LMs	31%	25%
Part of LM-Arousal pairs out of all Arousals	**57%**	**57%**
Part of PLM-Arousal pairs out of all PLMs	28%	24%
Part of PLM-Arousal pairs out of all Arousals	**42%**	**43%**
Part of iLM-Arousal pairs out of all iLMs	40%	30%
Part of iLM-Arousal pairs out of all Arousals	15%	14%

Table 14: Statistics for the occurrence of associated events for all patients together and differentiated in original and modified arousal annotations

Considering that 57% (original annotations and modified annotations) of all arousals, see Table 14, are associated with any kind of LM, it can be supposed that there is a relation between LM and arousals in patients with PLMS. Especially PLMs seem to be related to arousals, as 42% and 43% out of all arousals are associated with a PLM.

Although percentages of associated pairs seem to be too high to be random, a heuristic model is provided to support the hypothesis that the two events aren't independent from each other.

Let assume that all recordings are segmented in segments of 5s and that each segment is described by two random variables S^A (for arousals) and S^{LM} (for LMs). The binary variables S^A and S^{LM} take only two values within each segment $i = 1, \dots, L$ $(number\ of\ segments\ in\ all\ recordings)$:

$$S_i^A = \begin{cases} 1, & if\ segment\ i\ contains\ an\ arousal \\ 0, & if\ segment\ i\ doesn't\ contain\ an\ arousal \end{cases}$$

$$S_i^{LM} = \begin{cases} 1, & if\ segment\ i\ contains\ a\ LM \\ 0, & if\ segment\ i\ doesn't\ contain\ a\ LM \end{cases}$$

It is possible that one arousal or one leg movement can be part of more than one segment and therefore S_i^A and S_i^{LM} are not completely independent in time. Nevertheless, for the purpose of getting an approximate idea of how likely it is that an arousal and a LM appear simultaneously, it is assumed that S_i^A and S_i^{LM} are independently distributed.

Let the probability that a segment contains an arousal be $P(S_i^A = 1) = p$ and the probability that it contains a LM be $P(S_i^{LM} = 1) = q$.

In order to test if events are independent from each other the following hypothesis are proposed:

$$H0: P(S_i^A = 1, S_i^{LM} = 1) = p \cdot q, \qquad i = 1, \dots, L$$
$$H1: P(S_i^A = 1, S_i^{LM} = 1) > p \cdot q, \qquad i = 1, \dots, L$$

The null hypothesis is that the two events appear randomly and independent from each other. Or in other words: That the unconditional probability of the occurrence of a LM or an arousal is equal to the conditional probability of the occurrence of a LM given that the segment contains an arousal and vice versa. The values in Table 12 and Table 13 can be considered as an approximation for the conditional probabilities. To get an approximation for the unconditional probabilities p and q, the real number of 5s segments containing arousals and LMs is counted and divided by the total number of segments (when the patient is asleep and will be asleep for the next 30s).

	# segments with arousal	#segments with LM	# all segments in recording	% of all segments that contain an arousal	% of all segments that contain an LM
TS100714	468	535	3826	12%	14%
TS090714	472	707	4686	10%	15%
TC070814	182	149	5682	3%	3%
MB200814	159	342	5012	3%	7%
Sum/Mean	***1281***	***1733***	***19206***	***7%***	***10%***

Table 15: Approximation for the unconditional probability of the occurrence of an arousal or a LM

According to Table 15 the empirical approximations for p and q are on average:

$$p \approx 7\%, q \approx 10\%$$

Under the $H0$, that events are independent from each other, the conditional and the unconditional probabilities for both events should be the same. For instance, in the case of patient TS100714 and the modified arousal annotation set, the approximate unconditional probability of 12% for an arousal in Table 15 is compared with the approximate conditional probability of 29% in Table 13 that an arousal

occurs given that there is a LM. Equally, for the same patient and annotation set, the approximate unconditional probability for a LM of 14% in Table 15 is compared with the approximate conditional probability of 59% in Table 13 that a LM occurs given that there is an arousal. The conditional probability is in both cases above the unconditional probability. Doing so for all patients it can be observed, that the conditional probability approximation is always above the unconditional probability approximation.

Therefore, the heuristic model supports that the $H0$ should be rejected and the events are not independent from each other.

3.2.2 Analysis of cause/effect relation between LM and arousal events

This section also focuses exclusively on associated LM and arousal events. It is researched if there is any kind of causality between them. First, associated events are analyzed by calculating in which part of all pairs the arousal precedes the LM and vice versa.

	Original annotations	Modified annotations
Part of **LM-Arousal pairs** with **LM start first**	52%	56%
Part of **LM-Arousal pairs** with **Arousal start first**	48%	44%
Part of **PLM-Arousal** pairs with **PLM start first**	***53%***	***58%***
Part of **PLM-Arousal** pairs with **Arousal start first**	***47%***	***42%***
Part of **iLM-Arousal** pairs with **iLM start first**	50%	50%
Part of **iLM-Arousals** pairs with **Arousal start first**	50%	50%

Table 16: Part of associated events in which arousal precedes LM and vice versa for both arousal annotation sets

Table 16 shows that for all differentiations the LM precedes the arousal in more than 50% of the cases. Especially when considering PLM-Arousal pairs the percentage of PLMs preceding arousals even reaches 58%.

To test if these percentages are statistically significant above 50%, a binomial test can be performed:

Let the difference in start times within LM-Arousal pairs be modeled with a random variable X that takes only two values:

$$LM\ preceding\ Arousal = 1$$
$$Arousal\ preceding\ LM = 0$$

Let $X \sim B(n, p_0)$ and let assume that $p_0 = 0.5$ and $n = number\ of\ LM - Arousal\ pairs$. The hypothesis to be tested is:

$$H0: p = p_0$$
$$H1: p > p_0$$

Under the $H0$ the variable X follows a binomial distribution with probability p_0 for X equal to 1. If the $H0$ is rejected, the calculated percentage of LMs preceding arousals is significantly above 50%.

	Original annotations	Modified annotations
LM preceding arousal	0.38318923054	0.0258844600225
PLM preceding arousal	0.336389532122	0.0101416632378
iLM preceding arousal	1.0	1.0

Table 17: p-values for the binomial test if percentage of LM preceding arousals is significantly above 50%

All the p-values shown in Table 17 for original annotations are clearly above a significance level of $\alpha = 0.05$ and the $H0$ is accepted in all cases. For the modified annotations though, the p-values of "LM preceding arousal" and "PLM preceding arousal" indicate that the $H0$ is rejected with a significance level of $\alpha = 0.05$. That means that for this significance level LMs precede arousals significantly, which is partly caused by the fact that modified annotations were reviewed regarding precision in start and end time.

In order to test in another way if one event precedes more often the other, the difference in start times of associated events is calculated. As described in *section Analysis of cause/effect relation between LM and arousal events*, let $I = \{1, \dots, N\}$ be the index for the LM start times and $J = \{1, \dots, K\}$ the index for the arousal start times. Furthermore, let for LM associated with an arousal $L_i, i \in I_{LM_{pair}}$ be $A_{j(i)}$ the associated arousal start time.

$$L_i - A_{j(i)} > 0 \; if \; arousal \; precedes \; LM$$

$$L_i - A_{j(i)} < 0 \; if \; LM \; precedes \; arousal$$

Now the distribution of start differences is estimated with a histogram and displayed in Figure 43 and Figure 44.

Original annotations:

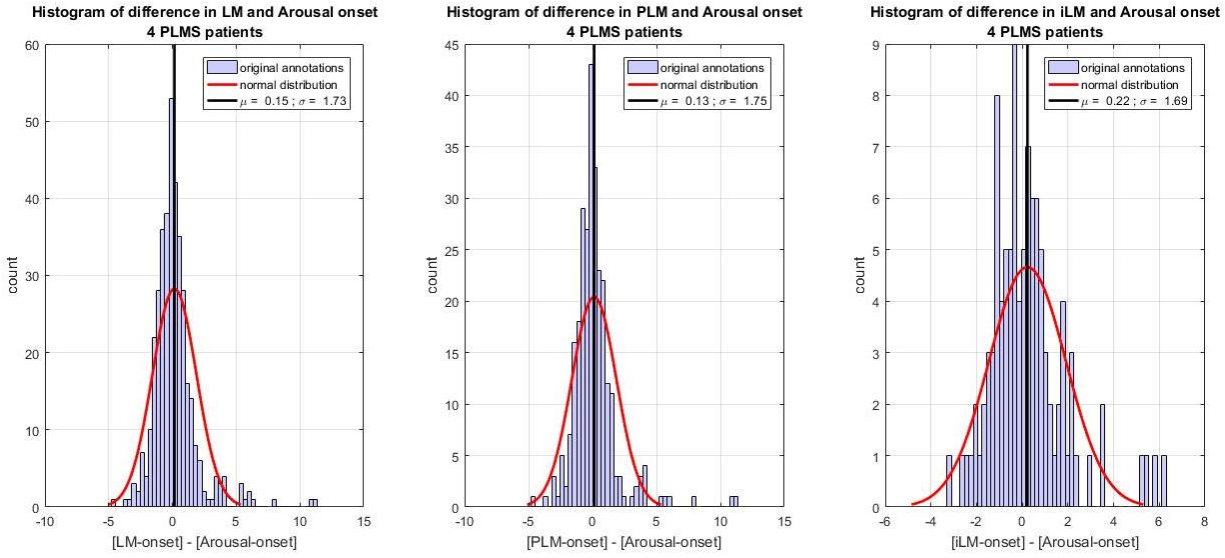


Figure 43: Difference in start times of all LM, PLM and iLM to their associated arousals. Here the original arousal annotation set is used. Positive values mean that the arousal precedes the LM.

Modified annotations:

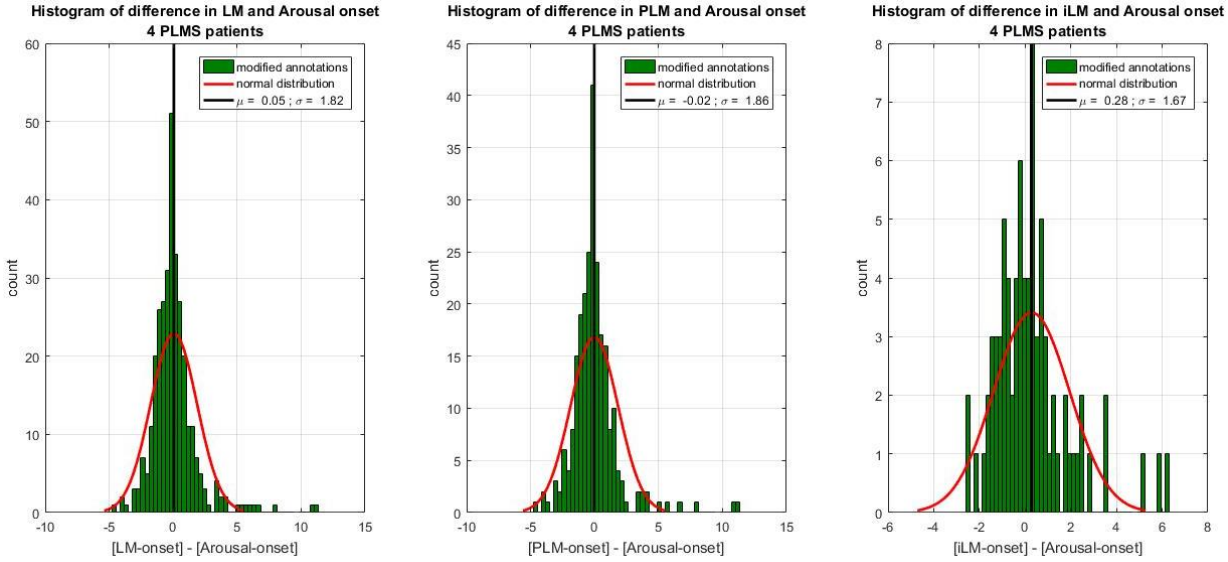


Figure 44: Difference in start times of all LM, PLM and iLM to their associated arousals. Here the modified arousal annotation set is used. Positive values mean that the arousal precedes the LM.

In Figure 43 and Figure 44 a normal distribution is fitted and an excess kurtosis as well as a potential longer right tail (arousals preceding LMs) can be obtained.

To test for normal distribution the Shapiro-Wilk-Test is applied:

$$H0: data\ is\ normally\ distributed$$
$$H1: data\ is\ not\ normally\ distributed$$

The following table displays all p-values for all types of leg movements and the two annotation sets:

	Original annotations	Modified annotations
[LM onset – Arousal onset]	2.0177189207337552e-19	6.363591442237823e-18
[PLM onset – Arousal onset]	4.102200829477584e-18	1.4009898477806526e-16
[iLM onset – Arousal onset]	3.940061105822679e-06	6.8845793066429906e-06

Table 18: P-values of Shapiro Wilk test on normality of the distribution of the difference in start times

According to the p-values in Table 18, the $H0$ is rejected in all cases for all reasonable significance levels, which means that none of the start differences is normally distributed.

Considering that the right tail seems to be longer than the left tail, it could be interesting to test whether the mean of the data set is significantly greater than zero. A possible test statistic is a one-sided t-test with hypothesis:

$$H0: \mu = 0$$
$$H1: \mu > 0$$

	Original annotations	Modified annotations
[LM onset – Arousal onset]	4.229592e-02	2.978582e-01
[PLM onset – Arousal onset]	1.137213e-01	5.708760e-01
[iLM onset – Arousal onset]	9.009532e-02	6.727113e-02

Table 19: P-values of one sided t-test of mean equal to zero against mean greater than zero

At a significance level of $\alpha = 0.05$ the $H0$ is accepted in all cases for both annotation sets, with the exception of the LM-Arousal start difference for original annotations. At a significance level $\alpha = 0.01$ the $H0$ is accepted in all cases, which means that the mean of the difference in start times is not significantly different from zero.

3.2.3 Analysis of relation in duration between LM and arousal events

Motivated by the work of (Ferri, et al., 2015) the relation between the duration of associated LM and arousal events is investigated. The duration of an event might be interpreted as its "intensity". Therefore, the hypothesis is that more intense LMs cause more intense arousals. The duration of associated events is visualized in Figure 45 and Figure 46 and a regression line is estimated with ordinary least squares to be able to tell something about the relationship between them. Let l_i be the duration of a leg movement and $a_{j(i)}$ the duration of its associated arousal, then the following model is estimated: $a_{j(i)} = C + \beta \cdot l_i + u_t$

Original annotations:

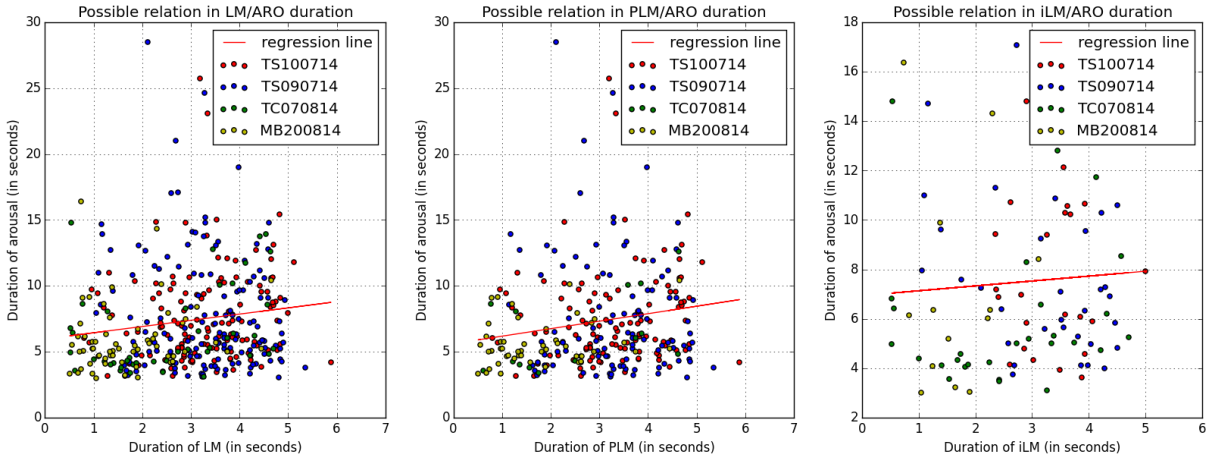


Figure 45: Scatterplot of duration of associated events and its resulting regression line for original arousal annotation set

	$\widehat{C}$	$\widehat{\beta}$
LM/Aro duration regression	5.96448804	0.4721265
PLM/Aro duration regression	5.60978248	0.56814512
iLM/Aro duration regression	6.93455087	0.19763298

Table 20: Estimated parameters C and beta for regression model: duration of arousals (with original annotations) regressed on duration of leg movements

Modified annotations:

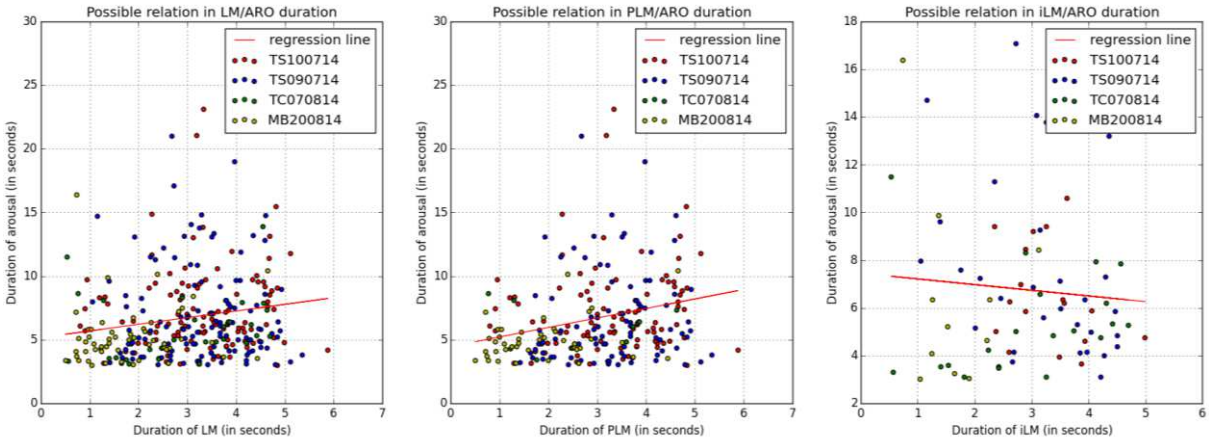


Figure 46: Scatterplot of duration of associated events and its resulting regression line for modified arousal annotation set.

	$\widehat{C}$	$\widehat{\beta}$
LM/Aro duration regression	5.16060486	0.52296987
PLM/Aro duration regression	4.46507626	0.74644042
iLM/Aro duration regression	7.45016716	- 0.2376719

Table 21: Estimated parameters C and beta for regression model: duration of arousals (with modified annotations) regressed on duration of leg movements

Analyzing these scatterplots graphically, it can be suspected that for these 4 patients there is no kind of relation between the duration of isolated LMs and arousals. In addition, it can be observed that the sign of the estimated coefficient $\hat{\beta}$ is positive for original annotations and negative for modified annotations, which also indicates that there is no clear relation in the duration for iLMs. Considering LMs and PLMs, the plotted regression lines show for both annotation sets a slightly positive slope. This might indicate a positive relation between the intensity of PLMs and arousals.

To test on the significance of the positive slope a t-test can be considered. To fulfill the assumptions of the t-test the residuals are tested on heteroscedasticity and further on normal distribution. For heteroscedasticity, the residuals are tested with the White-Test with hypothesis:

$$H0: homoscedasticity\ of\ residuals$$
$$H1: heteroscedasticity\ of\ residuals$$

If homoscedasticity is rejected one can use heteroscedasticity-robust standard errors (White estimation of variance).

	Original annotations	**Modified annotations**
LM/Aro dur. regression resid.	0.66786688326544053	0.15252011850954322
PLM/Aro dur. regression resid.	0.42354632428916661	0.061622300672801888
iLM/Aro dur. regression resid.	0.39091964962046166	0.16700708532355565

Table 22: P-values for white-test on heteroscedasticity of residuals of the regression of arousal duration on LM/PLM/iLM duration in associated pairs for both arousal annotations

According to Table 22, for a significance level of $\alpha = 0.05$ the $H0$ of homoscedasticity is accepted in all cases.

Furthermore, the normality of the residuals is tested with a Shapiro-Wilk test:

$$H0: residuals\ are\ normally\ distributed$$
$$H1: residuals\ are\ not\ normally\ distributed$$

	Original annotations	Modified annotations
LM/Aro dur. regression resid.	6.193013358244434e-20	1.9654601311744926e-18
PLM/Aro dur. regression resid.	3.295008622836594e-18	5.530128293670781e-16
iLM/Aro dur. regression resid.	6.239530421225936e-07	1.9579608760977862e-06

Table 23: P-values for test on normality of residuals of the regression of arousal duration on LM/PLM/iLM duration in associated pairs for both arousal annotations

As provided in Table 23 the $H0$ of normality is rejected in all cases.

So the residuals are homoscedastic but not normally distributed. There is no need for a robust variance estimator and the t-test statistic for the significance of the slope of the regression can be considered, with the assumption that the sample size (for LMs and PLMs between 243 and 400 samples; for iLMs between 80 and 108 samples) is large enough that asymptotical results for the classical regression model hold (see Table 14).

	Original annotations	Modified annotations
LM/Aro dur. regression resid.	5.15674176e-03	2.08059687e-03
PLM/Aro dur. regression resid.	4.81658311e-03	1.41562414e-04
iLM/Aro dur. regression resid.	5.21864112e-01	4.83156634e-01

Table 24: P-values for t-test on significance of slope coefficient of the regression of arousal duration on LM/PLM/iLM duration in associated pairs for both arousal annotations

Table 24 shows the significance of the positive slopes for all LMs and for PLMs. The p-value of the test for the coefficient in the regression considering only iLMs (for both annotation sets) is above a significance level of $\alpha = 0.05$ and therefore the $H0$ that the slope-coefficient is equal to zero is accepted. In other words, there is no relation between the duration of iLMs and the duration of their associated arousals, but there is a slightly positive relation in the duration of PLMs and the duration of their associated arousals.

To be sure that the above investigated results hold within the heterogenic data, consisting of 4 different patients, the regression model was estimated again for each patient and all LMs (dividing into subsets of PLMs and iLMs per patient would result in even smaller samples). It has to be noted that the sample size within each patient is between 62 and 149 samples (see Table 10) for original annotations and between 31 and 120 samples (see Table 11) for modified annotations, and therefore relatively small. That's why asymptotical results will possibly not hold for these regression models and the p-value for the t-test statistic on the significance of the coefficients is just provided for information.

Original annotations: **Modified annotations:**

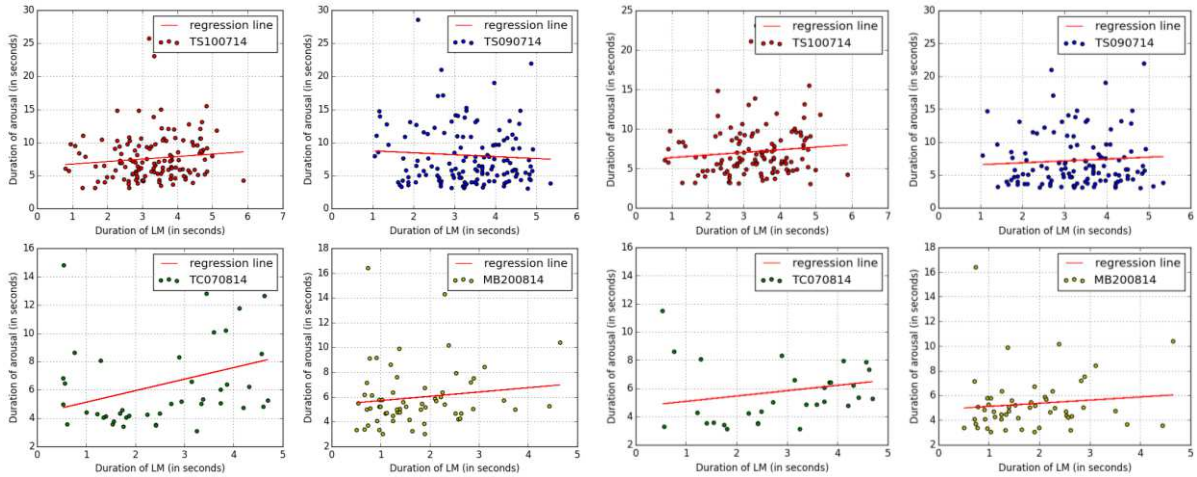


Figure 47: Relation in duration between the two events for each patient

Original annotations:

LM/Aro duration regression:	$\widehat{C}$ (p-value)	$\widehat{\beta}$ (p-value)
TS100714	6.346015 (8.432642e-09)	0.379013 (2.074548e-01)
TS090714	9.022891 (1.725058e-09)	-0.291764 (4.696231e-01)
TC070814	4.320019 (0.000125)	0.813125 (0.023596)
MB200814	5.363859 (4.213682e-11)	0.346396 (3.108069e-01)

Table 25: P-values for t-test on significance of the coefficients of the regression per patient (original arousal annotations)

According to Table 25 none of the informative p-values for the slope coefficient is significant, with the exception of patient TC070814. As already mentioned, this can be partly caused by the fact that the sample size is relatively small. But with exception of patient TS090714 coefficients are positive and relatively close to the estimated value $\hat{\beta} \approx 0.47$ for the whole set of patients.

Modified annotations:

LM/Aro duration regression	$\widehat{C}$ (p-value)	$\widehat{\beta}$ (p-value)
TS100714	6.042321 (1.172367e-08)	0.330424 (2.493219e-01)
TS090714	6.283738 (0.000015)	0.284976 (0.480324)
TC070814	4.699030 (0.000199)	0.378725 (0.283418)
MB200814	4.850422 (6.356049e-09)	0.255183 (4.552524e-01)

Table 26: P-values for t-test on significance of the coefficients of the regression per patient (modified arousal annotations)

According to Table 26 none of the informative p-values for the slope coefficient is significant. As already mentioned, this can be partly caused by the fact that the sample size is relatively small. All coefficients are positive and relatively close to the estimated value $\hat{\beta} \approx 0.52$ for the whole set of patients.

All in all, the investigated slightly positive relation between the duration of LMs and arousals, is supported by all patients for the modified annotation set and by all but one patient for the original annotation set.

3.2.4 Analysis of relation between the intensity of a LM and the occurrence of an arousal

A slightly different approach and a clinical interesting question is if more intense LMs are more likely to be associated with arousals. This could be an important finding for the diagnosis of sleep disorders caused be LMs. If LMs aren't causing any disruption of sleep or other problems, they are less important for diagnosis. Important are the leg movements that disturb sleep, which can be measured in the appearance of arousals. The first question in this context is how to measure intensity of a leg movement. In the following two ideas for a measure of the intensity are modeled: the duration and the "area" value computed by the 3D detection.

Duration as a measure of the intensity of a LM

When considering the duration as a measure of the intensity of a LM the question is if longer leg movements are more likely to cause disruptions of sleep. Table 27 shows the mean of the duration of LMs associated with arousals and not associated with arousals.

	Original annotations	**Modified annotations**
Mean duration of **LMs not** associated **with Arousal**	2,38 s	2,41 s
Mean duration of **LMs** associated **with Arousal**	2,98 s	3,02 s
Mean duration of **PLMs not** associated **with Arousal**	**2,39 s**	**2,40 s**
Mean duration of **PLMs** associated **with Arousal**	**3,01 s**	**3,05 s**
Mean duration of **iLMs not** associated **with Arousal**	2,33 s	2,37 s
Mean duration of **iLMs** associated **with Arousal**	2,89 s	2,95 s

Table 27: Mean of the duration in seconds of LMs associated with arousals and not associated with arousals

Original annotations:

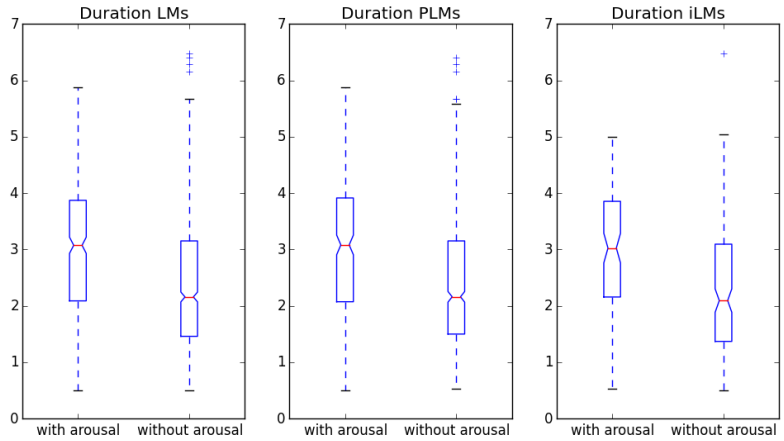


Figure 48: Boxplot comparison of the distribution of duration of LMs appearing with and without arousals for the original arousal annotation set

Modified annotations:

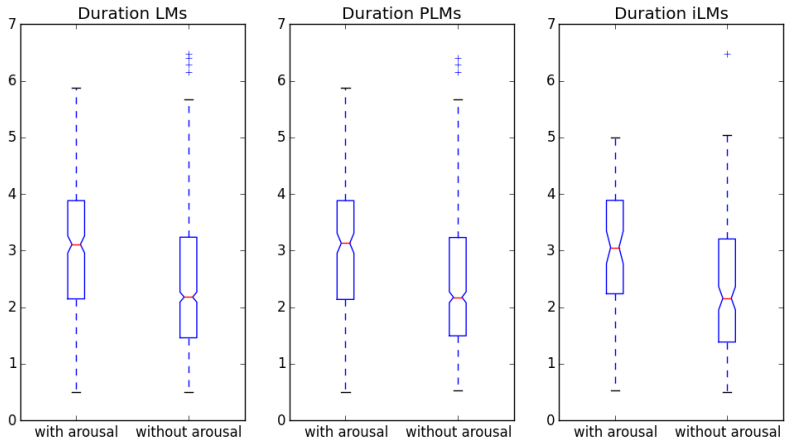


Figure 49: Boxplot comparison of the distribution of duration of LMs appearing with and without arousals for the modified arousal annotation set

According to Table 27, Figure 48 and Figure 49 LMs associated to arousals are on average longer than LMs that appear without arousal. The notch of the boxes within the plots do not overlap, which provides a graphical information about the significant difference in the median within the two groups. To test if they are statistically significantly longer than the ones without arousal another possibility is to perform a two sample t-test on the difference of the mean in the group of LMs with and the group of LMs without arousal.

Let μ_{LM+A} be the mean of the duration of LMs associated with an arousal and $\mu_{LM\setminus A}$ the mean of the duration of LMs without an associated arousal. (The same test is done exclusively for PLMs and iLMs)

$$H0: \mu_{LM+A} = \mu_{LM\setminus A}$$
$$H1: \mu_{LM+A} \neq \mu_{LM\setminus A}$$

	Original annotations	Modified annotations
LM with and without arousal	3.9889024494984853e-16	2.2008212058631565e-15
PLM with and without arousal	2.5504536090848689e-13	5.2205341633108542e-13
iLM with and without arousal	0.00021899827733267127	0.00035000000939217643

Table 28: P-values of two sample t-test on equal or not equal mean between two groups: LM/PLM/iLM with arousal and LM/PLM/iLM without arousal

The $H0$ is rejected in all cases for all reasonable significance levels. This means that the mean in the group associated with an arousal and the group not associated with an arousal are significantly different from each other.

This shows that for the used data set it holds that longer leg movements are more likely to be associated with an arousal and therefore to disturb sleep and cause sleep disorder problems for patients.

3D "Area" value as a measure of the intensity of a LM

As already described in the section *Analysis of relation between the intensity of a LM and the occurrence of an arousal*, in the part regarding the measure of intensity with the 3D detection, firstly the 3D detection itself is validated by calculating a set of numbers and percentages.

	Number of 3D LM detections	Number of PSG LM annotations	Number of PSG LM annotations -3D detected	Number of PSG LM annotations – not 3D detected
TS100714	1370	408	349 (86%)	59 (14%)
TS090714	1450	505	427 (85%)	78 (15%)
TC070814	106	118	30 (25%)	88 (75%)
MB200814	279	269	183 (68%)	86 (32%)
SUM/overall %	***3205***	***1300***	***989 (76%)***	***311 (24%)***

Table 29: Analysis of the performance of the 3D detection on the used patients

According to Table 29, 76% of the PSG annotated LMs within these 4 patients are detected by the 3D detection. In the following tables, it is shown how many LMs, separated in PLMs and iLMs, are associated with arousals. Doing so, it is again distinguished between the original and the modified annotation set for the arousals.

Original arousal annotations:

	3D detected					
	Number of LMs with arousals	**Number of LMs without arousals**	**Number of PLMs with arousals**	**Number of PLMs without arousals**	**Number of iLMs with arousals**	**Number of iLMs without arousals**
TS100714	127 (36%)	222 (64%)	103 (36%)	185 (64%)	24 (39%)	37 (61%)
TS090714	138 (32%)	289 (68%)	109 (32%)	237 (68%)	29 (36%)	52 (64%)
TC070814	24 (80%)	6 (20%)	8 (100%)	0 (0%)	16 (73%)	6 (27%)
MB200814	51 (28%)	132 (72%)	39 (24%)	124 (76%)	12 (60%)	8 (40%)
SUM/overall %	***340 (34%)***	***649 (66%)***	***259 (32%)***	***546 (68%)***	***81 (44%)***	***103 (56%)***

Table 30: Percentages and number for 3D detected LMs/PLMs/iLMs and their association to arousals (original annotation)

	Not 3D detected					
	Number of LMs with arousals	**Number of LMs without arousals**	**Number of PLMs with arousals**	**Number of PLMs without arousals**	**Number of iLMs with arousals**	**Number of iLMs without arousals**
TS100714	14 (24%)	45 (76%)	13 (24%)	41 (76%)	1 (20%)	4 (80%)
TS090714	8 (10%)	70 (90%)	1 (2%)	59 (98%)	7 (39%)	11 (61%)
TC070814	23 (26%)	65 (74%)	6 (21%)	23 (79%)	17 (29%)	42 (71%)
MB200814	11 (13%)	75 (87%)	10 (12%)	71 (88%)	1 (20%)	4 (80%)
SUM/overall %	***56 (18%)***	***255 (82%)***	***30 (13%)***	***194 (87%)***	***26 (30%)***	***61 (70%)***

Table 31: Percentages and number for not 3D detected LMs/PLMs/iLMs and their association to arousals (original annotation)

	Percentage of LMs with arousals that are not 3D detected	Percentage of PLMs with arousals that are not 3D detected	Percentage of iLMs with arousals that are not 3D detected
TS100714	10%	11%	4%
TS090714	5%	1%	19%
TC070814	49%	43%	52%
MB200814	18%	20%	8%
SUM/overall %	***14%***	***10%***	***24%***

Table 32: Percentages of LMs/PLMs/iLMs associated with arousals (original annotations) that are not 3D detected

Table 30 and Table 31 distinguish between 3D detected LMs and not 3D detected LMs. It is of high interest to know how many and what percentage of LMs associated with arousals are 3D detected and not 3D detected. As shown in Table 30 and Table 31 the majority of LMs aren't associated with an arousal, but according to Table 32 only 14% out of the LMs with arousals are dismissed by 3D detection. Considering only the LMs that are not 3D detected, Table 31 shows that 82% of them are not associated with any arousal. As arousals can be seen as short disruptions of sleep, which can cause various problems, like daytime sleepiness, the LMs associated with arousals may have a higher influence on sleep quality. This investigation shows that 3D detection is able to detect 86% of all LMs with arousals among these 4 patients.

Modified arousal annotations:

3D detected						
	Number of LMs with arousals	**Number of LMs without arousals**	**Number of PLMs with arousals**	**Number of PLMs without arousals**	**Number of iLMs with arousals**	**Number of iLMs without arousals**
TS100714	116 (33%)	233 (67%)	97 (34%)	191 (66%)	19 (31%)	42 (69%)
TS090714	120 (28%)	307 (72%)	95 (27%)	251 (73%)	25 (31%)	56 (69%)
TC070814	19 (63%)	11 (37%)	7 (88%)	1 (12%)	12 (55%)	10 (45%)
MB200814	47 (26%)	136 (74%)	36 (22%)	127 (78%)	11 (55%)	9 (45%)
SUM/overall %	***302 (31%)***	***687 (69%)***	***235 (29%)***	***570 (71%)***	***67 (36%)***	***117 (64%)***

Table 33: Percentages and number for 3D detected LMs/PLMs/iLMs and their association to arousals (modified annotations)

Not 3D detected						
	Number of LMs with arousals	**Number of LMs without arousals**	**Number of PLMs with arousals**	**Number of PLMs without arousals**	**Number of iLMs with arousals**	**Number of iLMs without arousals**
TS100714	10 (17%)	49 (83%)	10 (19%)	44 (81%)	0 (0%)	5 (100%)
TS090714	5 (6%)	73 (94%)	0 (0%)	60 (100%)	5 (28%)	13 (72%)
TC070814	14 (16%)	74 (84%)	3 (10%)	26 (90%)	11 (19%)	48 (81%)
MB200814	7 (8%)	79 (92%)	7 (9%)	74 (91%)	0 (0%)	5 (100%)
SUM/overall %	***36 (12%)***	***275 (88%)***	***20 (9%)***	***204 (91%)***	***16 (18%)***	***71 (82%)***

Table 34: Percentages and number for not 3D detected LMs/PLMs/iLMs and their association to arousals (modified annotations)

	Percentage of LMs with arousals that are not 3D detected	Percentage of PLMs with arousals that are not 3D detected	Percentage of iLMs with arousals that are not 3D detected
TS100714	8%	9%	0%
TS090714	4%	0%	17%
TC070814	42%	30%	48%
MB200814	13%	16%	0%
SUM/overall %	***11%***	***8%***	***19%***

Table 35: Percentages of LMs/PLMs/iLMs associated with arousals (modified annotations) that are not 3D detected

As already displayed for original annotations the Table 33 and the Table 34 distinguish between 3D detected LMs and not 3D detected LMs. The results for the modified annotation set remain the same as for the original annotation set in its message. As shown in Table 33 and Table 34 the majority of LMs aren't associated with an arousal, but according to Table 35 only 11% out of the LMs with arousals are dismissed by the 3D detection. Considering only the LMs that are not 3D detected, Table 34 shows that 88% of them are not associated with any arousal. As arousals can be seen as short disruptions of sleep, which can cause various problems, like daytime sleepiness, the LMs associated with arousals may have a higher influence on sleep quality. This investigation shows that 3D detection is able to detect 89% of all LMs with arousals among these 4 patients.

The following figures show the difference in distribution of the area value within the groups with and without arousals. The LMs that couldn't be detected by 3D detection are assigned with the area value zero and LMs that are detected by more than one 3D detection are assigned with the sum of the area values of all overlapping 3D detection.

Original arousal annotations:

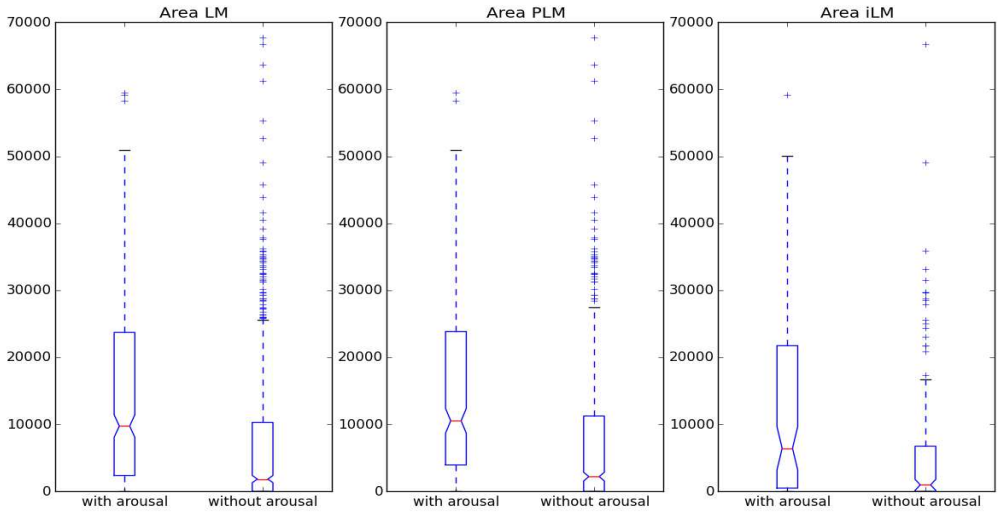


Figure 50: Boxplot of distribution of area values within the groups of LMs/PLMs/ILMs associated and not associated with arousals (original annotations)

Modified arousal annotations:

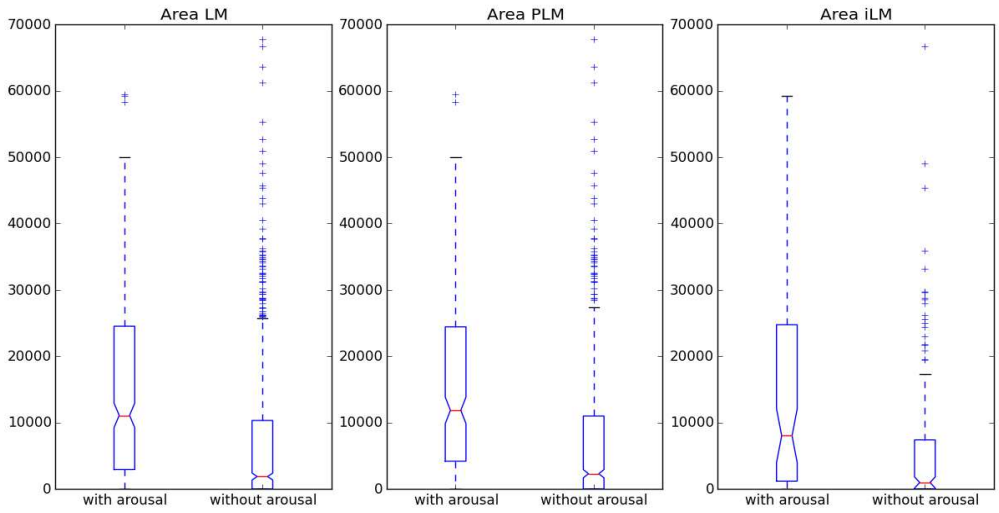


Figure 51: Boxplot of distribution of "area" values within the groups of LMs/PLMs/ILMs associated and not associated with arousals (modified annotations)

Figure 50 and Figure 51 show that the median (notches of boxes aren't overlapping) is significantly different between the groups with and without arousals. In addition, it can be tested with a t-test if the means are significantly different from each other. Let μ_{LM+A} be the mean of the area value of LMs associated with an arousal and $\mu_{LM\setminus A}$ the mean of the area value of LMs without an associated arousal. (The same test is done exclusively for PLMs and iLMs)

$$H0: \mu_{LM+A} = \mu_{LM\setminus A}$$
$$H1: \mu_{LM+A} \neq \mu_{LM\setminus A}$$

	Original annotations	Modified annotations
LM with and without arousal	1.2510805741058368e-16	9.6608002573776594e-20
PLM with and without arousal	3.9498716690715905e-15	1.3375796879087781e-16
iLM with and without arousal	0.001877574074490275	9.8550283782640761e-05

Table 36: T-test on difference in mean of "area" values between the groups with and without arousals for LMs/PLMs/iLMs and both arousal annotation sets

According to Table 36 all mean values are significantly different from each other because $H0$ is rejected in all cases until a significance level of $\alpha = 0.001$.

Next to the approach of visualizing the distribution with boxplots one can think of categorizing area values in 4 categories of intensity and visualizing in which category the most LMs with or without arousals appear. Figure 52 shows the distribution of the area value for 3D detected LMs. The values are divided into four categories according to the quantiles of the distribution in Figure 52 plus one category zero for LMs that are not 3D detected.

Quantiles of the distribution:

- $q_{0.25}^{3D} = 2006{,}77;\ q_{0.5}^{3D} = 8134{,}82;\ q_{0.75}^{3D} = 18113{,}51$

Categories 0-4:

- Category = 0 if area == 0 (not 3D detected LM)
- Category = 1 if area ∈ $(0, q_{0.25}^{3D})$
- Category = 2 if area ∈ $[q_{0.25}^{3D}, q_{0.5}^{3D})$
- Category = 3 if area ∈ $[q_{0.5}^{3D}, q_{0.75}^{3D})$
- Category = 4 if area ∈ $[q_{0.75}^{3D}, \max(area)]$

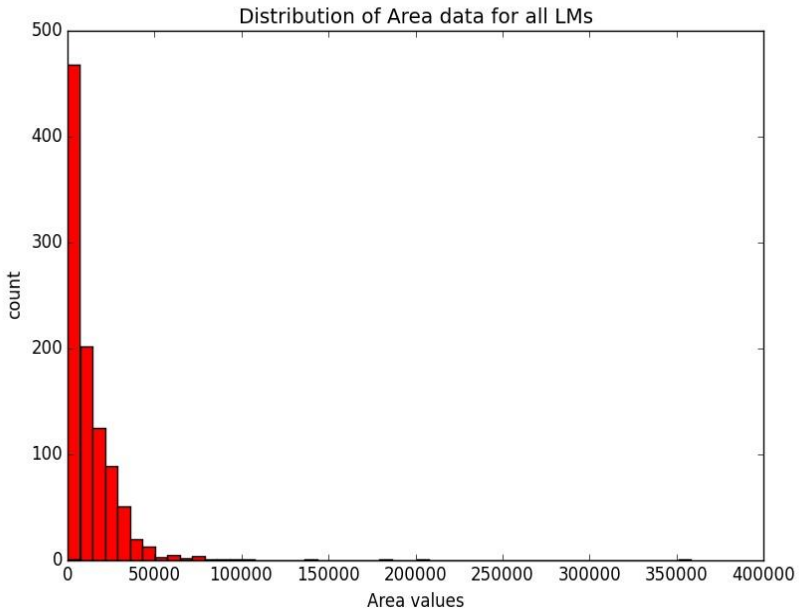


Figure 52: Distribution of area values for 3D detected LMs. Zero values for the not 3D detected LMs are not included here

Figure 53 and Figure 54 show the distribution of the intensity of the LMs in the above described categories. It can be observed that the distribution of LMs with arousals is right-skewed and the one for LMs without arousals left-skewed. That means that most LMs with arousals have intensity values of category 4 and that most LMs without arousals are in category zero, which means that they aren't even detected by 3D detection because no movement can be recognized and it's maybe only a muscle tension.

This supports the hypothesis that more intense arousals are more likely to occur with arousals and therefore could be more important for the diagnosis of sleep disorders. It also supports the 3D detection performance because the not detected LMs are about 50% less likely to occur with an arousal and therefore less important in terms of sleep disruption.

Original annotations:

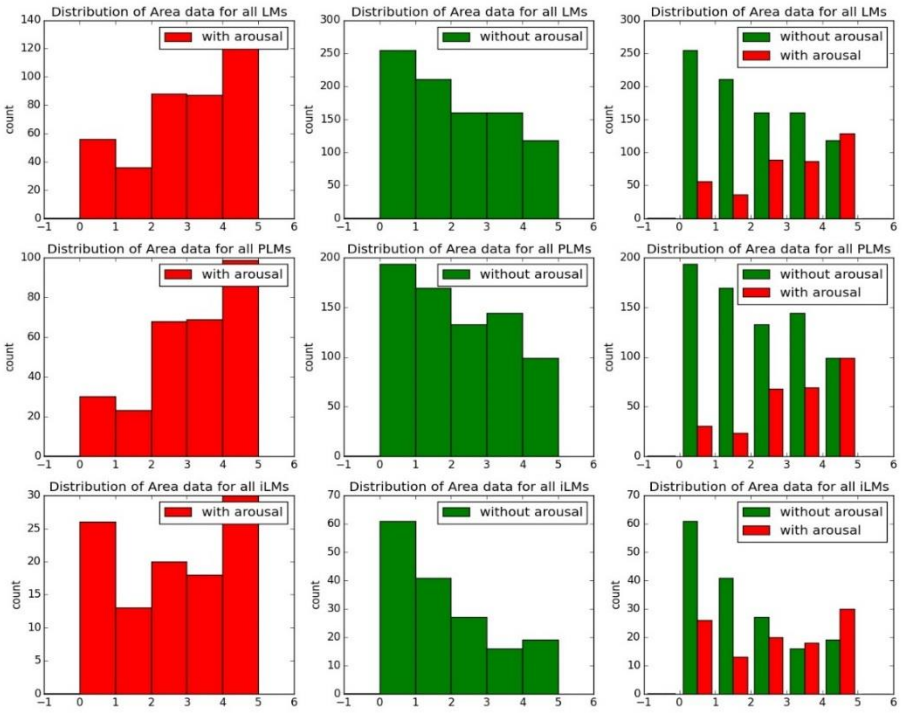


Figure 53: Histogram of the 4 categories of intensity in area values and the zero assignments (for not 3D detected LMs) separated in the group with and without arousals (original annotations)

Modified annotations:

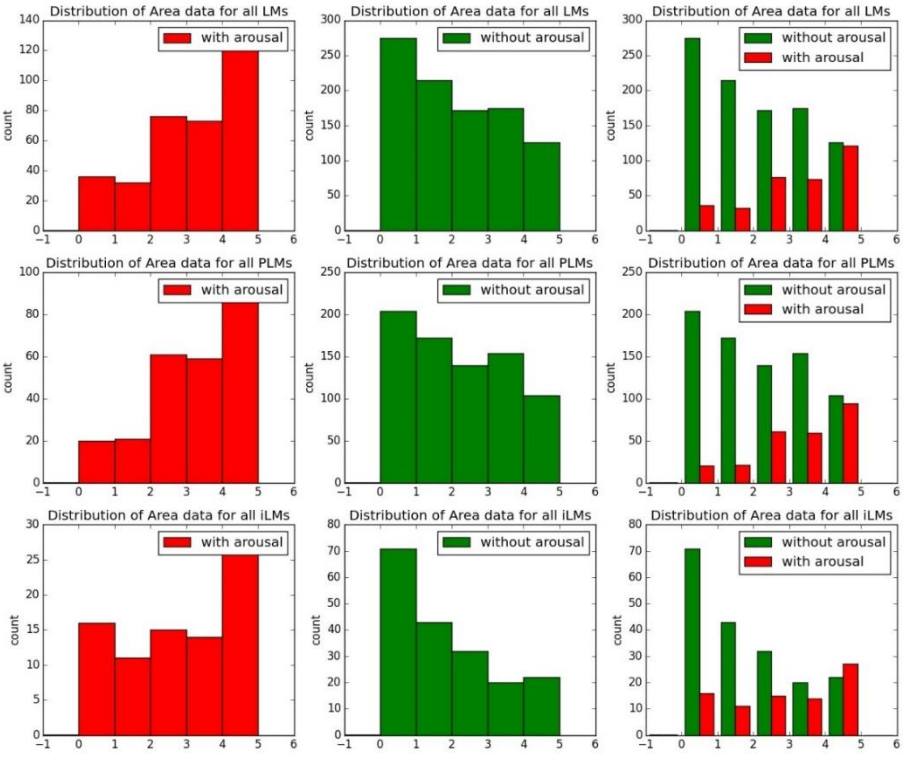


Figure 54: Histogram of the 4 categories of intensity in area values and the zero assignments (for not 3D detected LMs) separated in the group with and without arousals (original annotations)

4 Discussion and Conclusion

The investigations in this thesis relate on the PSG data and a manually revised set of arousal annotations of a group of seven patients of the AKH Wien. In the first part of this thesis the development of an algorithm for the automatic detection of arousals in EEG-signal was investigated. Due to the definition of the AASM of an arousal the method consists of two main steps: the detection of the start of an arousal and the final determination of an arousal.

More precisely, the whole night recording for each patient was divided in segments of 3s and features were extracted for each segment. The first step of the classification of the segments as arousal start segments, described by a set of 13 features (per EEG-channel), extracted exclusively from two central EEG derivations, was realized with a Support Vector Machine with radial basis kernel. New findings in feature extraction were the usage of the coefficients of an AR(6)-model, estimated within each segment for each EEG channel, and the usage of a t-test statistic on a significant difference in the mean of the squared filtered signal. The parameters for the Support Vector Machine were chosen by performing a grid search with Leave-One-Out-Cross validation for a set of parameters and the decision was made by calculating common statistical performance values and considering the ROC-Curve as well as the Youden-Index. With the chosen parameters $C = 11$ and $\gamma = 2^{-12}$ arousal start segments could be detected with an average sensitivity and specificity of about 96% and a positive predictive value of about 24%.

The second step and final determination of the arousals was done by calculating the duration of each arousal and by passing the set of detections through several algorithms to check for arousal criteria like minimum length of 3s, simultaneous change in submental EMG for REM-sleep arousals and 10s of stable sleep before the start of an arousal. Due to the fact that neither the end of an arousal, nor the 10s of stable preceding an arousal are defined properly in the definition of the AASM, the duration of an arousal and the validity of the 10s stable sleep are rather subjective and performance of the detection algorithm dropped in terms of sensitivity but it improved in terms of selectivity after the second step. After determining final arousals, the average sensitivity was about 86% and the positive predictive value reached on average 60%. Compared to other works on the automatic arousal detection, described in the state of knowledge, these were still satisfying results. In particular, further investigations on the second step of the detection algorithm as well as on a precise definition of the end of an arousal and the preceding stable sleep, could improve the final performance.

The second aim of this thesis was to investigate the relation between different kinds of leg movements and arousals. It was supported that leg movements and arousals can't be completely independent of each other but a clear causality couldn't be defined, even though it was tested that periodic leg movements precede arousals slightly significantly more often (for the modified annotation set). Furthermore, the analysis on the relation between the intensity of a leg movement and the association to an arousal led to interesting new findings. Two measures for intensity, the duration and an intensity value of the 3D detection of leg movements, were tested and it could be shown that more intense leg movements are more likely to occur with an arousal. Considering that arousals are kind of short disturbances of sleep, that means that more intense leg movements can be more important for the diagnosis of sleep disorders. In addition, the performance of the 3D detection was supported by demonstrating that the number of leg movements associated with an arousal is 50% less for not 3D

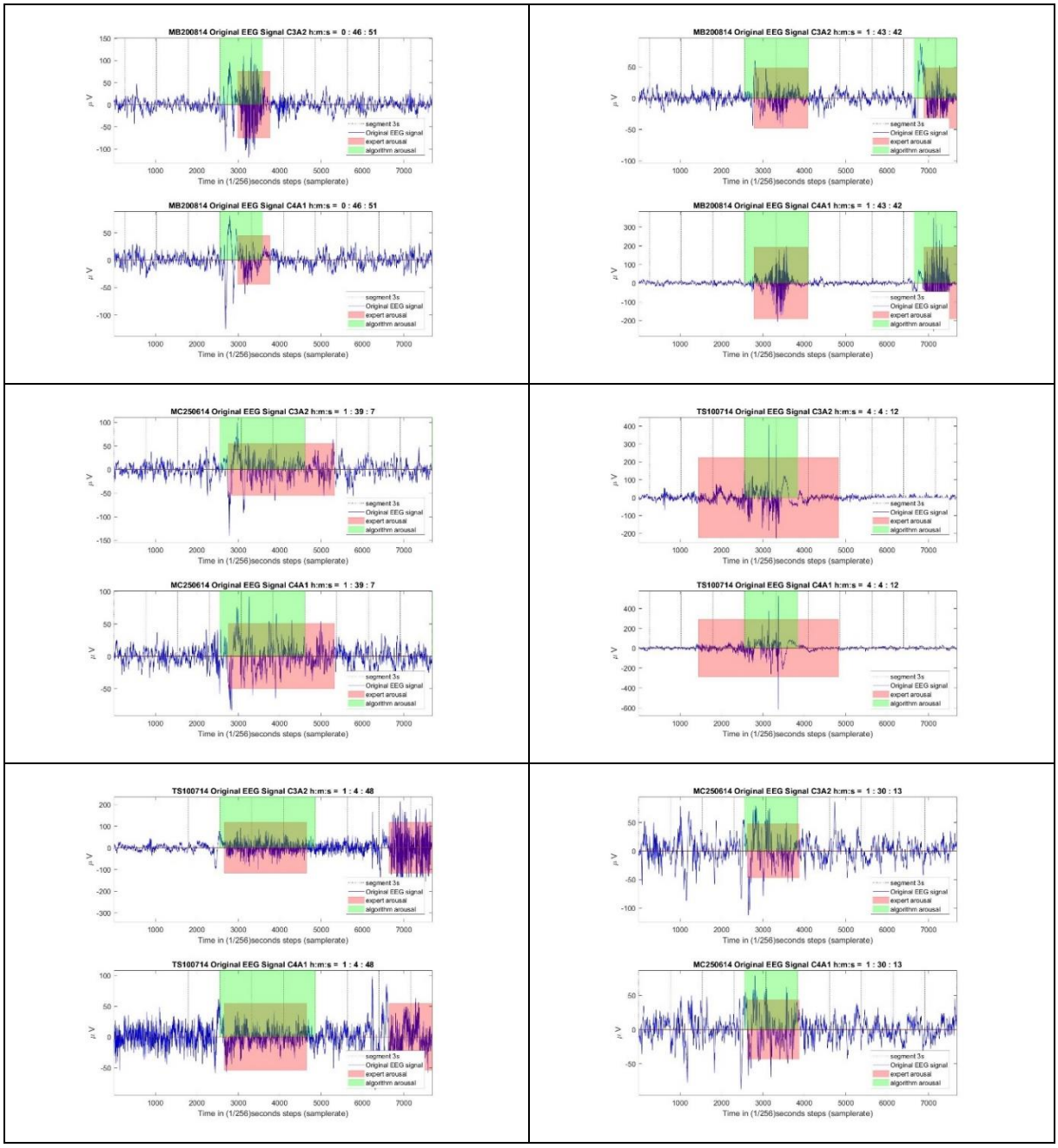
detected leg movements than for the ones that got 3D detected. Hence, the great majority of not 3D detected leg movements is not associated with an arousal and therefore less important in terms of sleep disruptions.

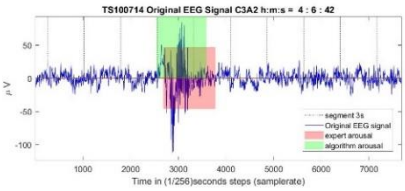
For further research and after the findings in the relation between leg movements and arousals it could be interesting to investigate on a new set of features for the arousal detection that consider the EMG and the 3D leg signals. It can be supposed that these features could be especially suitable for PLMS patients and could lead to a higher performance in arousal detection.

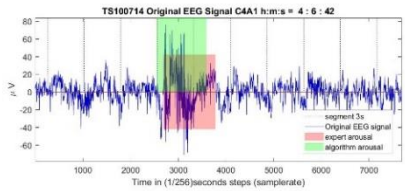
5 Appendix

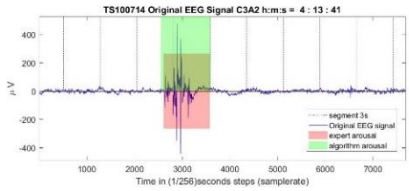
In addition to chapter *Results of the arousal detection,* there are more plots of automatic arousal detections provided in this section.

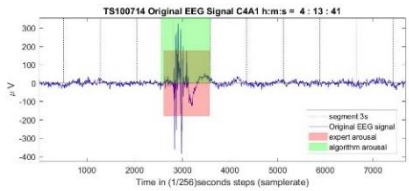
True Positives:



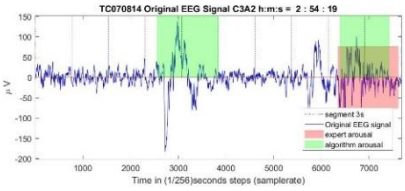


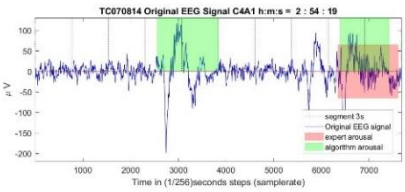


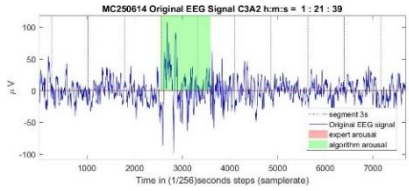


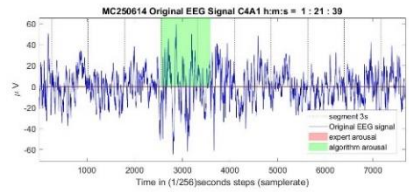


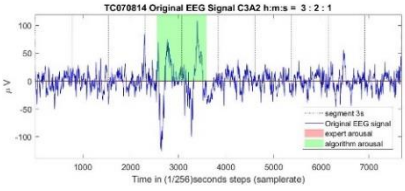
False Positives:

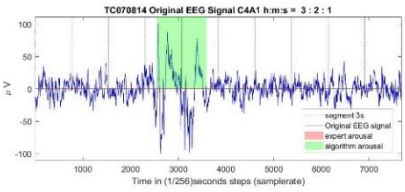


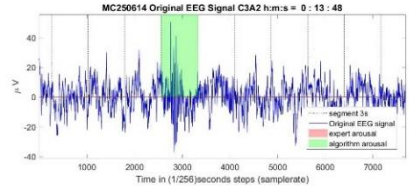


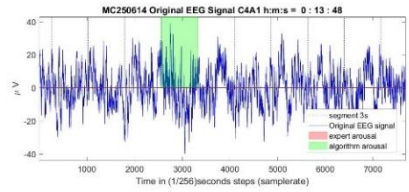


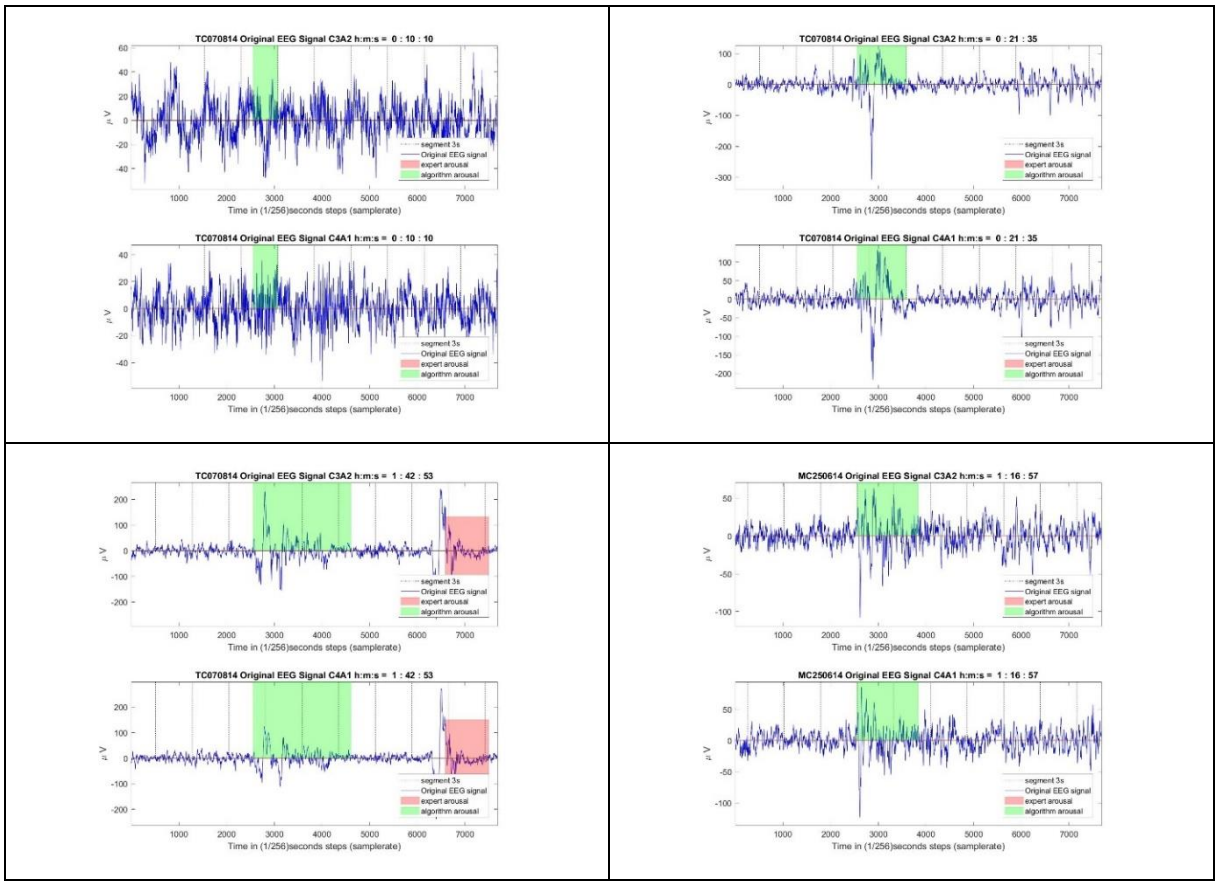




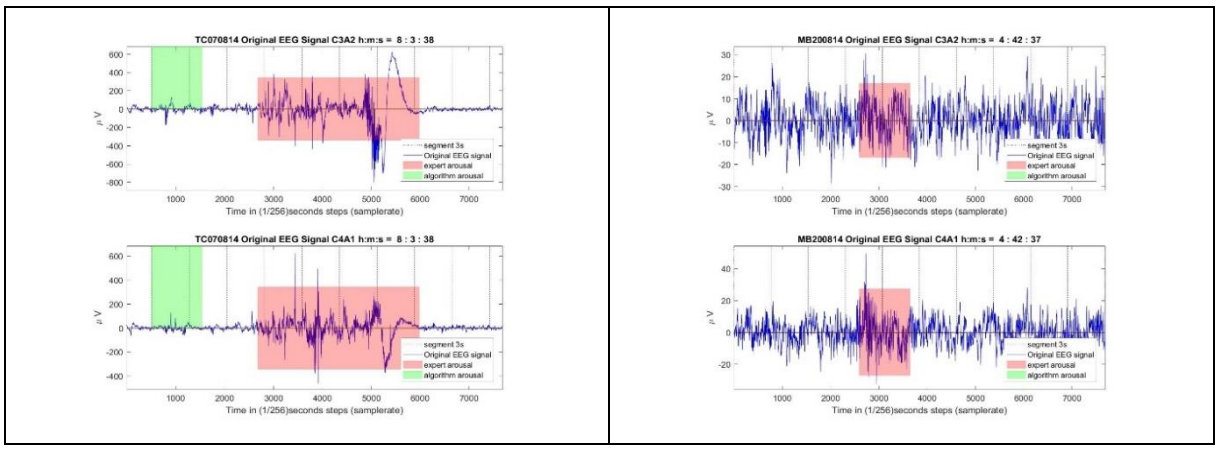


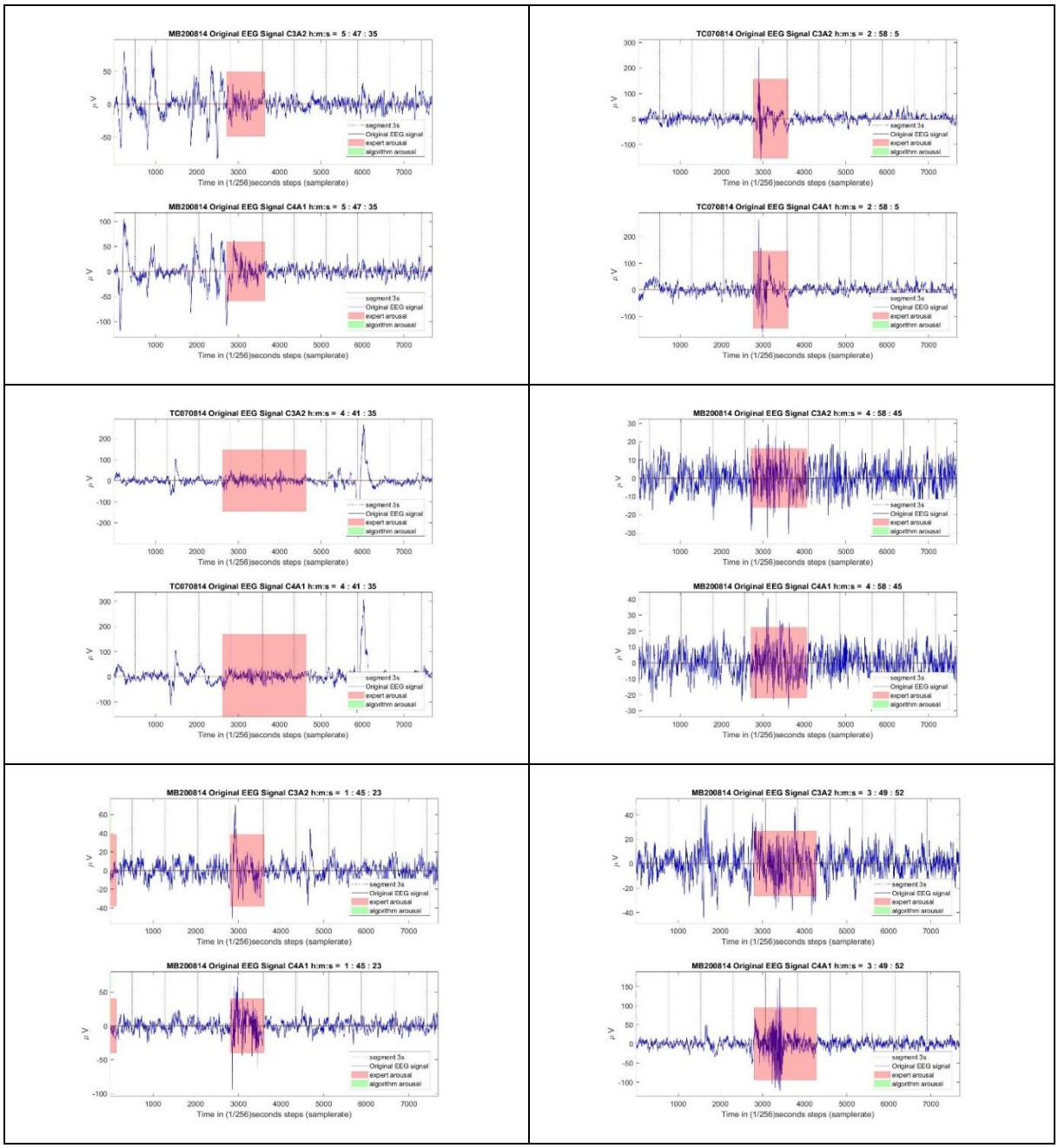






False Negatives:





6 Abbreviations

ACF	**A**uto**c**orrelation **F**unction
ANN	**A**rtificial **N**eural **N**etwork
ECG	**E**lectro**c**ardio**g**raphy
EEG	**E**lectro**e**ncephalo**g**raphy/**E**lectro**e**ncephalo**g**ram
EMG	**E**lectro**m**yo**g**raphy/**E**lectro**m**yo**g**ram
NREM	**N**on**r**apid **E**ye **M**ovement
PACF	**P**artial **A**uto**c**orrelation **F**unction
LM	**L**eg **M**ovement
iLM	**i**solated **L**imb **M**ovement
PLM	**P**eriodic **L**imb **M**ovement
PLMS	**P**eriodic **L**imb **M**ovement of **S**leep
PLMD	**P**eriodic **L**imb **M**ovement **D**isorder
PSG	**P**oly**s**omno**g**raphy/**P**oly**s**omno**g**ram
REM	**R**apid **E**ye **M**ovement
SVM	**S**upport **V**ector **M**achine
TST	**T**otal **S**leep **T**ime
AASM	**A**merican **A**cademy of **S**leep **M**edicine

7 List of Figures

8 Bibliography

Agarwal, R. (1-4. September 2005). *Automatic Detection of Micro-Arousals.* Shanghai, China: IEEE.

Álvarez-Estévez, D., & Moret-Bonillo, V. (2009). Model Comparison for the Detection of EEG Arousals in Sleep Apnea Patients. *IWANN 2009*, (S. 997-1004).

Álvarez-Estévez, D., & Moret-Bonillo, V. (1. January 2011). Identification of Electroencephalographic Arousals in Multichannel Sleep Recordings. *IEEE Transactions on Biomedical Engineering*(58), S. 54-63.

Berry, R. B., Brooks, R., Gamaldo, C. E., Harding, S. M., Lloyd, R. M., Marcus, C. L., & Vaughn, B. V. (2016). *The AASM Manual for the Scoring of Sleep and Associated Events.* (A. A. Medicine, Hrsg.)

Carney MD, P. R., Geyer MD, J. D., & Sachin PHD, T. (2012). Introduction to Sleep and Polysomnography. In P. R. Carney MD, J. D. Geyer MD, & T. Payne MD, *Atlas of Polysomnograpy* (S. 258-280). Lippincott WIlliams & Wilkins.

Chaudhary, B. A. (2007). Introduction to Polysomnography. In B. A. Chaudhary, MD, FCCP, FACP, & FAASM, *Primary Care Sleep Medicine* (S. 295-304). Totowa, NJ: Humana Press Inc.

Cho, S., Lee, J., Park, H., & Lee, K. (2005). Detection of Arousals in Patients with Respiratory Sleep Disorders Using a Single Channel EEG. *Engineering in Medicine and Biology*, (S. 2733-2735). Shanghai, China.

De Carli, F., Nobili, L., Gelcich, P., & Ferrilo, F. (Feb 1999). A Method for the Automatic Detection of Arousals During Sleep. *SLEEP, No. 5*(Vol. 22), 561-572.

Ferri, R., Rundo, F., Zucconi, M., Manconi, M., Bruni, O., Ferini-Strambi, L., & Fulda, S. (2015). An Evidence-based Analysis of the Association between Periodic Leg Movements during Sleep and Arousals in Restless Legs Syndrom. *SLEEP*(Vol. 38, No. 6), 919-924.

Guido, S., & Müller, A. C. (2016). Kernelized Support Vector Machines. In S. Guido, & A. C. Müller, *Introduction to Machine Learning with Python: A Guide for Data Scientists* (S. 94-106).

Hornyak, M., Feige, B., Voderholzer, U., & Riemann, D. (2005). Spectral analysis of sleep EEG in patients with restless legs syndrom. *Clinical Neurophysiology*(116), 1265-1272.

medistat, G. (2016). *Medistat*. Retrieved Feb 7, 2017, from Medistat: https://www.medistat.de/glossar/diagnostischer-test/youden-index/

Oppenheim, A. V., Schafer, R. W., & Buck, J. R. (2004). Filterentwurfstechniken. In A. V. Oppenheim, R. W. Schafer, & J. R. Buck, *Zeitdiskrete Signalverarbeitung* (S. 534-547). Pearson Studium.

Scherrer, W. (2015). Stationäre Prozesse und Zeitreihenanalyse. *Script for lecture*. Technische Universität Wien.

scikit-learn, d. (2010-2016). *scikit learn documentation 0.18.1*. Von http://scikit-learn.org/stable/modules/svm.html abgerufen

Shahrbabaki, S. S., Dissanayaka, C., Patti, C. R., & Cvetkovic, D. (2015). *Automatic Detecion of Sleep Arousal Events from Polysomnographic Biosignals.* Melbourne, Australia: IEEE.

Sorensen, G. L., Kempfner, J., Jennum, P., & Sorensen, H. B. (2011). Detection of arousals in Parkinson's disease patients. *33rd Annual International Conference of the IEEE EMBS* (S. 2764-2767). Boston, Massachusetts, USA: IEEE.

Sugi, T., Kawana, F., & Nakamura, M. (2008). Automatic EEG Arousals Detection for Obstructive Sleep Apnea Syndrome. *The International Federation of Automatic Control* (S. 5209-5214). Seoul, Korea: IFAC.

Sugi, T., Kawana, F., & Nakamura, M. (2009). Automatic EEG arousal detection for sleep apnea syndrome. *Biomedical Signal Processing and Control*, 329-337.

Trans Cranial Technologies ldt. (29. December 2016). Von Trans-Cranial-Technologies: https://www.trans-cranial.com/local/manuals/10_20_pos_man_v1_0_pdf.pdf abgerufen

Wallant, D., Muto, V., Gaggioni, G., Jaspar, M., Chellappa, S. L., Meyer, C., . . . Phillips, C. (2016). Automatic artifacts and arousals detection in whole-night sleep EEG recordings. (Elsevier, Hrsg.) *Journal of Neuroscience Methods*(258), 124-133.

Zhang, Y., Ji, X., & Zhang, Y. (2015). *Classification of EEG signals based on AR model and approximate entropy.* Liaoning Normal University, School of Computer and Information Technology. Dalian, China: IEEE.