Masteroppgave

NTNU
Norges teknisk-naturvitenskapelige universitet
Fakultet for informasjonsteknologi og elektroteknikk
Institutt for teknisk kybernetikk

Susann Masdal

A semiautomatic spindle detection platform with minimal expert intervention based on interaction between two software

Masteroppgave i Kybernetikk og robotikk
Veileder: Marta Molinas

Desember 2022

NTNU
Kunnskap for en bedre verden

Susann Masdal

A semiautomatic spindle detection platform with minimal expert intervention based on interaction between two software

Masteroppgave i Kybernetikk og robotikk
Veileder: Marta Molinas
Desember 2022

Norges teknisk-naturvitenskapelige universitet
Fakultet for informasjonsteknologi og elektroteknikk
Institutt for teknisk kybernetikk

NTNU
Kunnskap for en bedre verden

NTNU
Kunnskap for en bedre verden

Department of Engineering Cybernetics

TTK4900 - Teknisk kybernetikk, Masteroppgave

A semiautomatic spindle detection platform with minimal expert intervention based on interaction between two software

Author:
Susann Masdal

Supervisor:
Prof. Marta Molinas

December, 2022

Abstract

This study aims to find a more practical way to exploit two different public software detectors as a platform to minimize workload performed by sleep experts to annotate sleep spindles. The goal is for a laboratory or research center to be able to build up a customized database using their own database, where the sleep expert will spend the time verifying annotations instead of tedious and repetitive manual work for entire datasets. By using their own database, the target is for the machine learning software to still be able to detect spindles in data that include artifacts specific for the respective lab. Sleep spindles are a defining characterisation of N2 sleep in Non-Rapid Eye Movement (NREM) sleep, although these also occur in the other stages of NREM sleep.

Throughout the project, two different software detectors were chosen due to their distinct properties were the combination of these two eliminates the need for a sleep expert to identify these events. The public software detectors were the A7-algorithm (#A7) and Dreem One Shot Event Detector (DOSED), which specializes in event annotations. Sleep scoring was already available for all datasets used in this project, and therefore no automatic detectors were applied for this. The motivation is to remove time-consuming and monotonous work for a sleep expert, where the work put in place to use the proposed platform will free up time as well as provide more accurate annotations over time. There will be a somewhat more moderate workload in building the database, but with a larger and more diverse basis, the annotations will only require a person to start running a script while it predicts without the need for supervision.

The semiautomatic spindle detection platform shows potential where the workload for the sleep expert is drastically decreased. However will a bigger dataset be needed to fully emulate manual annotations. The pipeline would be applicable to detect other events, where the usability of DOSED can be further exploited since it is able to detect for multiple micro-architecture events.

The data were collected from the International Institute for Integrative Sleep Medicine (WPI-IIIS), University of Tsukuba, Japan. The data was gathered from three different studies. The cap used in the experiments consisted of 128 EEG channels, used to monitor the brain wave activity through electrodes attached to the scalp and face.

Sammendrag

Denne studien tar sikte på å finne en mer praktisk måte å utnytte to forskjellige offentlige programvaredetektorer som en plattform for å minimere arbeidsbelastningen utført av søvneksperter for å annotere søvnspindler ("sleep spindles"). Målet er at et laboratorium eller forskningssenter skal kunne bygge opp en tilpasset database ved hjelp av egen data, hvor søvneksperten skal bruke tiden sin på å verifisere annoteringer i stedet for kjedelig og repeterende manuelt arbeid. Ved å bruke sin egen database er målet at maskinlæringsprogramvaren fortsatt skal kunne oppdage spindler i data som inkluderer artefakter spesifikke for det respektive laboratoriet. Søvnspindle er en definerende karakteristikk for N2 søvn i Non-REM-søvn (NREM), selv om disse også forekommer i de andre stadiene av NREM.

Gjennom hele prosjektet ble to forskjellige programvaredetektorer valgt på grunn av deres distinkte egenskaper, der kombinasjonen av disse to eliminerer behovet for en søvn ekspert for å identifisere spindler i en søvnsyklus. De offentlige programvaredetektorene var A7-algoritmen (#A7) og "Dreem One Shot Event Detektor" (DOSED), som spesialiserer seg på annotering av ulike hendelser. Søvnskåring var allerede tilgjengelig for alle datasett som ble brukt i dette prosjektet, og derfor ble det ikke brukt automatisk detektorer for dette. Motivasjonen er å fjerne tidkrevende og monotont arbeid for en søvnekspert, hvor arbeidet som legges til grunn for å bruke de foreslåtte kombinerte detektorene vil frigjøre tid samt gi mer nøyaktige merknader. Det vil være noe mer moderat arbeidsbelastning i det man bygger opp databasen, men med et større og mer mangfoldig grunnlag vil annoteringen kun kreve at en person begynner å kjøre programvaren mens den predikerer uten behov for tilsyn.

Den semiautomatiske spindeldeteksjonsplattformen viser potensiale der arbeidsmengden til søvnekspertene reduseres drastisk. Derimot vil et større datasett være nødvendig for å bedre etterligne manuelle merknader. "Pipeline"-modellen vil være anvendelig for å oppdage andre hendelser, der brukervennligheten til DOSED kan utnyttes ytterligere siden den er i stand til å detektere for flere mikroarkitekturhendelser.

Dataen ble samlet inn fra "International Institute for Integrative Sleep Medicine" (WPI-IIIS), Universitetet i Tsukuba, Japan. Dataen ble samlet inn fra tre forskjellige studier. Hetten som ble brukt i eksperimentene besto av 128 EEG-kanaler, brukt til å overvåke hjernebølgeaktivitetene gjennom elektroder festet til hodebunnen og ansiktet.

Preface

This Master's thesis is submitted in partial fulfillment of the five year Master's degree program Cybernetics and Robotics at the Norwegian University of Science and Technology (NTNU).

I would like to properly thank my supervisor, Professor Marta Molinas, for all help throughout the whole project, as well as her support and advice for the past couple of months. I will forever by grateful for this experience and your motivational words. I would also like to thank the Human Sleep lab of the International Institute of Integrative Sleep Medicine in Japan for providing the dataset, as well as guidance and motivation to learn more about sleep medicine during my time at their lab. A special thank you goes out to Professor Takashi Abe, for the warm welcome into his lab, as well as Ms. Yoko Suzuki for her cooperation on the PSG recordings and spindle annotations.

An honourable mention will also go out to the Sasakawa foundation for the scholarship that made it possible for me to travel to the Sleep lab in Japan for two months, to experience all stages involved when working at a sleep lab. Karine Lacourse provided me with useful information regarding her code for spindle detection, and I am very thankful for her contribution. Throughout the project I have utilized several Python packages made possible by all the maintainers keeping them up-to-date, as well as the IDUN cluster where some computations were made. A last thank you goes out to all sleep researchers that have worked with sleep classification and spindle annotation. Without their research and publications this would not have been possible.

The findings of this thesis were submitted to SLEEP 2023, the 37th annual meeting of APSS. A draft of the abstract can be seen in the Appendix.

Table of Contents

Chapter 1

Introduction

This chapter consist of four sections describing the motivation behind this project, as well as the background and short on previous work in this field. Information about the scope of the project as well as the structure of the report can be found at the end.

1.1 Background and motivation

Sleep is essential for maintaining a good health, as well as reduce risks caused by sleep deprivation. Scientist agree of the importance on sleep as well as all the benefits related to a good night of rest, while the purpose of sleep is somewhat unclear. Human spends approximately one-third of their life sleeping, where the recommended cycle lasts for about 7-9 hours depending on age [39]. During the sleep cycle information and memory from the day is stored, at the same time as waste fluid that surrounds the brain cells are flushed out and into the blood [33]. To be able to maintain a healthy body the cells need to be repaired, molecules like hormones and proteins are released, as well as energy needs to be restored. This is all done while our eyes are shut, and we enter the world of dreams. The nerve cells needs to be able to communicate and reorganize to support healthy brain functions. For most people the effects of a bad night of sleep is presented with emotional instability and non-functional brain capacity. The sleep cycle can be divided into five different stages depending on the brain activity, with two overall definitions for either light or deep sleep.

To be able to record sleep, and more specifically brain activity, Electroencephalograph (EEG) are used with small electrodes attached to the scalp. These electrodes monitor the electrical impulses the brain cells use to communicate. The use of this technique is also used when diagnosing or treating brain disorder, brain dysfunction, stroke, epilepsy as well as sleep disorder [58]. When a person is awake, several parts of the brain will be active to be able to perform simple as well as complex tasks. With the use of EEG one has found that many of the same areas of the brain also lights up when sleeping. The first experiments of these electrical activities where conducted on animals with the use of alternating lights. The fluctuating brain activity led to the conclusion of brain waves, and the experiment was presented in 1875 by Richard Caton [19]. However was the first EEG recording of a human obtained by Hans Berger in 1924.

The field of sleep research has previously considered the gold standard method to be manually scoring on recordings of the brain by visual inspection according to manuals describing different characteristics and stages that may occur during the sleep cycle. However has the development of machine learning and the use of several automatic detectors proven to be just as accurate as the annotations made by experts. The limitations for using machine learning is the requirement of a large and diverse dataset for a model to predict above a certain accuracy. This is needed to be able to outsource the manual scoring. Restrictions limits the sharing of data, and therefore makes it more difficult to build up dataset used in training.

1.2 Previous work

Writing "automatic spindle detector" in the Google search engine results in 5 630 000 results within 0.41 seconds. Most of these results are links to papers and projects comparing either the accuracy of a model against another or against manual annotations. New methods are developed from previous flawed models, or created to seize other difficulties that have risen due to the improvement in the field. However, as mentioned in the paper from Christian O'Reilly and Tore Nielsen [48],

> *" [...] new detectors can no longer be supported merely by threshold-dependent variables such as sensitivity and precision are superior to those of previously published detectors."*

There are many detectors already developed, and tested with a satisfying outcome. There is no need to continue developing new techniques as long as there is no more data to test on to verify that the model is sustainable and usable for several environments. The obstacle with automatic detection is that there is no agreed testing condition, as well as no record of how different conditions may interfere with the result when used by other research teams. The previously trustworthy results from sleep experts have been challenged by the development of detectors, and when EEG databases are difficult to acquire will the outcome be dependent on characteristics like size, diversity and reliability. There is no gold standard method to cover the obstacles, while there is a need to better utilize what is already created to minimize vulnerable interference.

To be able to detect a sleep spindle, both sleep experts and algorithms rely on the definitions given by either the R&M manual, or the modified guideline from American Academy of Sleep Medicine (AASM). The R&M manual was written by Allan Rechtschaffen and Anthony Kales in 1968, based on observations from young, healthy adults. The manual did therefore not apply to elderly subjects, or people with for instance sleep disorders. This was the motivation behind the modified AASM, which also included more detailed descriptions to limit the subjective interpretation as well as new directions for scoring methods. They wanted to confine the uncertainty so that this could not impact the scoring. The latest updated version of the AASM is from 2020, based on recommendations from the AASM Scoring Manual Committee [1].

1.3 Scope of this project

This report will focus on how to take advantage of public software detectors that emulate human experts, and will with time eliminate the monotonous work of manual annotation. Dataset from IIIS is run through an Independent Component Analysis to remove artifacts, before it is used in the A7-detector to annotate spindles with only sleep scoring and data as required information. The computations that results in spindle observations are then combined with the raw EEG data to found the basis for the other software detector, Dreem One Shot Event Detector (DOSED). This algorithm requires annotations and data to be able to run, but processing is done within the method. With time will this be the only detector needed to identify spindles, were several unknown datasets can run at the same time. The sleep scorer will only verify the result from the DOSED, where 10% of the data would give a good overview of the detectors performance.

1.4 Structure of the report

Chapter 2 includes relevant background theory regarding sleep, with focus on the different sleep stages and how electrodes monitor brain activity, in addition to different sleep classification techniques. The next chapter describes the methods used to remove artifacts, as well as spindle annotation using two software detectors. Subsequently is chapter 4 describing the results for the spindle detectors, and how one will with time eliminate tedious workload and predict more accurately events. The last chapter contains the summary and conclusion of the proposed method, as well as a discussion and mentions of further work.

Chapter 2

Theory

This chapter includes detailed information about sleep, as well as the four deviations and their characteristics. The explanation of EEG and how the measurements are done is also mentioned, including how current classification of sleep recordings are conducted. The two software detectors used in this report is also introduced at the end of the chapter.

2.1 Sleep

Throughout our life, sleep plays an essential role where getting enough hours of sleep is vital for maintaining a good health. Nevertheless, how the hours are divided into states and sleep stages are as important to make sure the quality of sleep is as restorative as requested.

Sleep can be categorized into to basic states of sleep, mainly Rapid Eye Movement (REM) and Non-rapid Eye Movement (NREM). Sleep spindles normally occurs in the NREM state, which is a state that again is divided into three different stages named N1, N2 and N3. Each stage represent a unique and perceptible brain wave pattern, where a sleep experts score a 30-second epoch based on the majority stage in the window. Most of the sleep for a human being is during NREM which is considered to be important for memory and rest. NREM is also known as the quiet sleep [11]. In 2007 the American Academy of Sleep Medicine (AASM) updated their classification of sleep, where they went from five sleep stages to four stages [18]. Before this, N4 was considered to be an separate stage, but is now included in the definition of stage N3. To illustrate the different sleep stages and what is considered to be a normal sleep cycle see figure (2.1).

As a humans falls asleep, they will during the earliest phases of sleep be relatively awake and alert. With the Electroencephalogram (EEG) attached to the scalp one may see the brain produce beta waves, which indicates that the brain is still active and engaged with small and fast brainwaves. Following this, as the brain is more relaxed and slows down, alpha waves would be present. Some experience strange and vivid sensations during this transition, where this phenomenon known as hypnagogic hallucinations [6]. This is often perceived as the feeling of falling or that the person imagines there are other people in the room. The hallucinations are common, and teens, young adults and women are most likely to have these hallucinations. For some these episodes also may occur in the transition from slumber to wakefulness.

The first stage a person encounter when falling asleep is stage N1, and a person awakened from this stage may report that they have not slept. This stage is defined as the transition between awake and falling a sleep, and is therefore categorized as light sleep. Therefore humans are easily awakened in this period, which normally last up to seven minutes. During this stage the muscles begin to relax, and both the breathing and heartbeat is slowing down. Some experience twitching in the body as a result of to the muscles beginning to lose tension. The brain is still fairly active during this time period and will produce high amplitude theta waves which occurs mostly in the brain's frontal lobe [18].

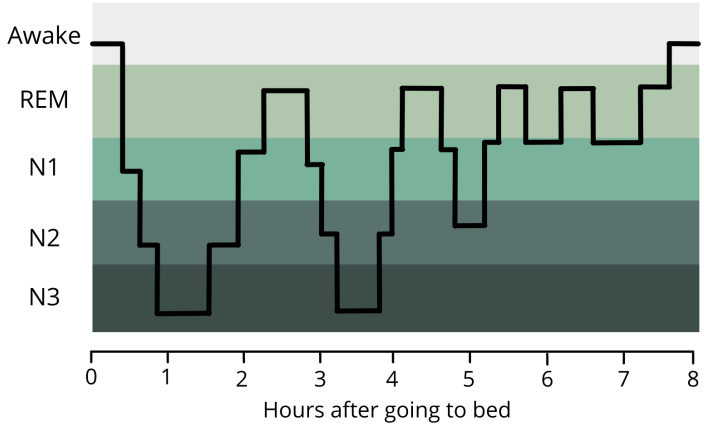


Figure 2.1: The different sleep stages for a typical 8-hour sleep cycle

The next stage is N2, which is a deeper stage than the previous one. The similarities is that this stage is also considered to be categorized as light sleep, and a person is still easily awakened. This is the stage where a person spends the majority of their sleep in, which is approximately 50% of the sleep time. Each stage last usually for about 10-25 minutes, and therefore occurs several times during the cycle. Entering this stage, the muscles relax even more, and there is a drop in body temperature, as well as slowed breathing and heartbeat. The eye movements will stop, and the brain produces burst of rapid, rhythmic brain activity known as sleep spindles. These are markers often associated with memory consolidation, which is when the brain gathers, processes and filters new memories from the previous day.

Following N2 comes stage N3, which is also called the slow wave sleep. The body becomes less responsive to the outside world, and it is harder to awaken a person. The stage lasts for about 20-40 minutes, where the period lasts longer at the first half of the sleep cycle, and will decrease with time. To be able to feel refreshed and re-energized, it is important to reach this stage, and the body starts is physical repairs. Delta waves will be present in the brain, which is often referred to as delta sleep. During this stage the brain consolidates declarative memories, which includes facts and statistics, general knowledge, personal experience or other things one may have learned [23].

REM sleep is the last stage, and occur approximately 90 min after a person initially falls asleep. This is the stage where people experience the most dreaming, and the muscles becomes temporarily paralyzed as to not reenact the dreams. The stage will during the sleep cycle increase in length, and normally lasts for about five minutes the first time. Due to the length increasing during the sleep, many experience to wake up during a dream, as the heart beat and blood pressure increases. The name of the cycle refers to the flickering of the eyes underneath the eye lids, and some also refer to this stage as "paradoxical sleep" since the brain activity is similar to when a person is awake. Memory consolidation also appear during REM sleep, and is believed that the emotional memories are processed and stored. This is also an important stage for learning since the brain uses time to cement information into memory.

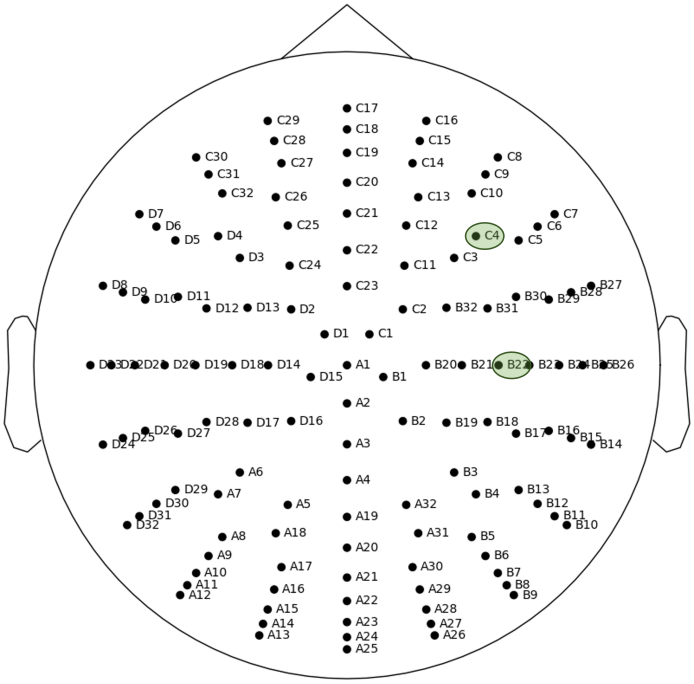


Figure 2.2: Biosemi128 with channel C4 and B22 indicated by the green circles, which were later used for spindle annotation

2.2 Data set

2.2.1 IIIS dataset

The data set was collected by the Human Sleep Lab of the International Institute for Integrative Sleep Medicine (IIIS), Tsukuba, Japan [28]. The data relevant for this project was collected from eight healthy, female participants, and was scored according to the AASM standard and routine Polysomnography (PSG) scoring. The age of the participants were 22.8 $\pm$ 2.0 [18-27] years (mean $\pm$ standard deviation [range]) [66]. The PSG recordings consist of 128 EEG channels, two mastoid channels, three Electromyographic (EMG) channels and three Electrooculogram (EOG) channels. The sampling frequency for all recordings were 1024Hz. The EEG channels were set up according to the biosemi128 configuration and according to fig (2.2) [27], and were scored based on the AASM manual with 30-second epochs [2]. The EOG and EMG were placed around the ear and face to detect muscle movements. In total 136 channels were used, recorded by Polysmith software [31] and PSG-1100 [32].

The scoring was done with sampling frequency set to 256Hz, while EEG and EOG was filtered with a low-frequency and high-frequency filter set to 0.3 and 35Hz. A filtering from 10 to 100Hz was applied to EMG. The sleep stages were split into five discrete stages according to AASM.

The data was collected from three different studies, either from test of procedural memory, observational study of menstrual cycle or a portable sleep EEG device validation study based on simultaneously measurements using PSG and EEG device. Participants with prior history of NREM parasomnias were excluded from this study, with a written consent gathered from the relevant participants. The International Review Board later approved the collection of data and usage (Approved ID#H29-177).

In addition to the PSG recordings, sleep scoring for all participants as well as spindle annotation for two subjects were completed by the sleep expert from the lab.

2.3 Brain functions

The outermost layer of the brain is called the cerebral cortex, and this is what gives the characteristic wrinkly appearance. Furthermore the cerebral cortex is divided lengthwise into two cerebral hemisphere which again is connected by a thick bundle of nerve fibres called the corpus callosum [25]. This is how the two sides are able to communicate and send signals to each other. Each hemisphere is traditionally divided into four lobes known as the frontal, parietal, temporal and occipital lobe [26]. The main organs of the Central Nervous System (CNS) is the spinal cord and brain, which again is the most complex organ of the body.

The frontal lobe is located at the front part of the brain, just behind the forehead. Due to it being the largest lobe in the human brain, it is also a common region of injury in traumatic brain injury. As seen in the figure (2.3), the frontal lobe is indicated by the red color, where the paired frontal lobes are known as either the left or right frontal lobe. The frontal lobe is associated with higher-level cognitive functions such as emotional regulation, planning, problem solving and reasoning. This is also why damage to this area may lead to personality changes, which is also the first sign of frontotemporal dementia [61].

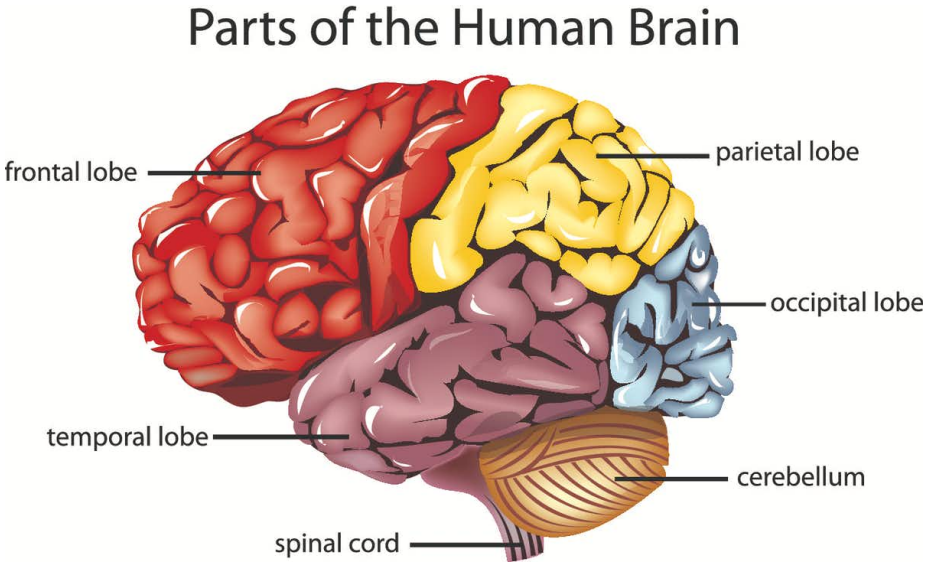


Figure 2.3: Illustration of the different parts of the brain seen from the side (Image source: [24])

The parietal lobe is located behind the frontal lobe, closer to the back and top of the head, indicated by the yellow color in figure (2.3). This area is responsible for integrating sensory information such as touch, temperature, pressure and pain. Damage to this area may affect a persons ability to finding their way around new or familiar places, as well as troubles with understanding spoken and/or written language [4].

At the far back of the head is the location to the occipital lobe, which is indicated by the light blue color in the figure above. This is the smallest lobe of the four, and is mainly responsible for the visual information from the eyes. This includes being able to perceive depth, color, distance and location of a seen object. Damage to this are may lead to visual deficits, for instance in terms of shapes and colors.

The last lobe is located behind the ares and is called the temporal lobe. This is the second largest lobe of the brain, and is indicated by the purple color. Processing of sensory information such as hearing, recognising language and forming memories is done in this area. Some areas also works on visual information to make sense of faces and scenes, in addition to the auditory information.

2.4 Neural activity

The human body would not be able to operate without the nervous system, which is responsible for the complex network that coordinates our actions, reflexes and sensations. The nervous system consists of two main parts, where the CNS is the most important in this case as it consists of the brain and spinal cord [55]. The name refers to the brains ability to receive information, as well as coordinate and influence activity in all parts of the human body. The CNS consists of nerve cells and glia cells, which are located between neurons. An illustration of how each nerve cell is build up can be seen in fig. (2.4).

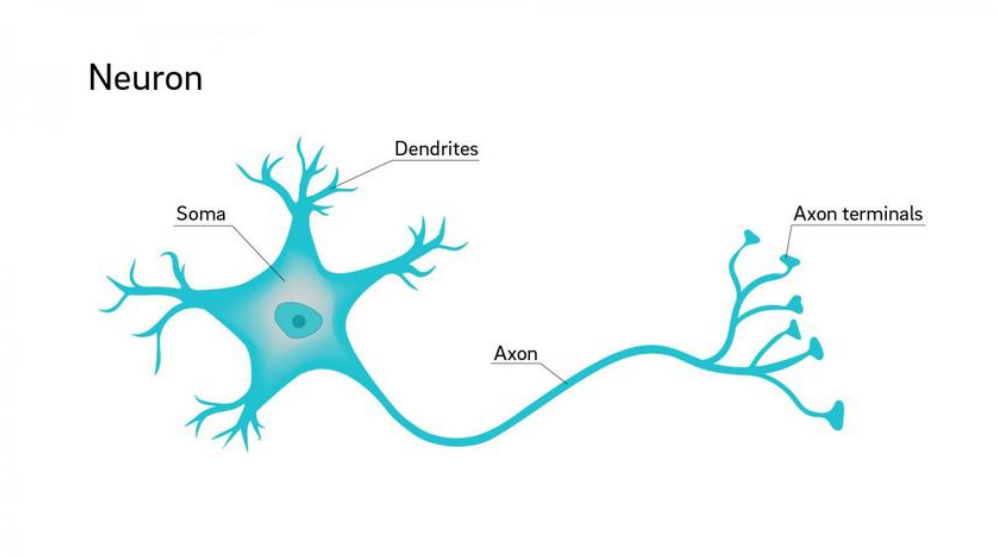


Figure 2.4: Illustration of a neuron (Image source: [15])

The connection of dendrites is either to the axons or dendrites to other cells, where they receive impulses from nerves or relay the signals to other nerves. The connection of one nerve is approximately made to 10.000 other nerves, where this is mostly done through dendrites connections. The main function of the axons are to transmit electrical impulses away from the cell, where the speed of the transmission is related to the diameter of the axon.

When the brain is in a resting stage, one can record a negative potential of approximately 60-70 mV under the membrane of the cell. A generated pulse will stimulate the potential, and is known as Action Potential (AP). This is only possible if there is a change in voltage above a set threshold over a short time period. Very weak stimuli will not produce a transmitted AP, but cause a small electrical disturbance. AP are short pulses that lasts for approximately five to ten ms, and are located in the membrane potential as well as transmitted along the axon.

2.5 Electroencephalogram

EEG is an electrophysical, non-invasive, monitoring method used to record activity in the scalp. To be able to detect the activity in the brain, small electrodes consisting of small metal disc with thin wires are pasted onto the scalp, usually attached to a cap to make the placement of the electrodes easier if there are many electrodes [40]. The paste is used to make sure the electrodes are as close to the scalp as possible, in addition to reduce skin impedance [13]. EEG is a convenient tool when looking at brain signals due to its small size, how easy it is to use as well as the limited need for additional equipment.

Within the brain there are billions of cells where half of them are neurons while the rest help to facilitate the activity of the neurons. The neurons are densely interconnected via synapses, which act as gateways of inhibitory of excitatory activity. To be able to generate an electrical field strong

Layer	**Thickness** [cm]	**Resistance** [ω]
Scalp	0.2 - 0.5	300 - 400
Skull	0.3 - 0.7	10 - 25 k
Cortex	0.1 - 0.3	50 - 150

Table 2.1: The thickness and resistance of three layers of the brain

enough to spread through tissue, bone and skull, thousands of neurons need to fire in sync. The burst of a signal neuron is difficult to detect without being in direct contact with it. The electrical filed is measured on the head surface, where [65] details that 10^7 neurons needs to be activated at the same time to get through all of the layers of the head, where an illustration of these layers are given by fig. (2.5).

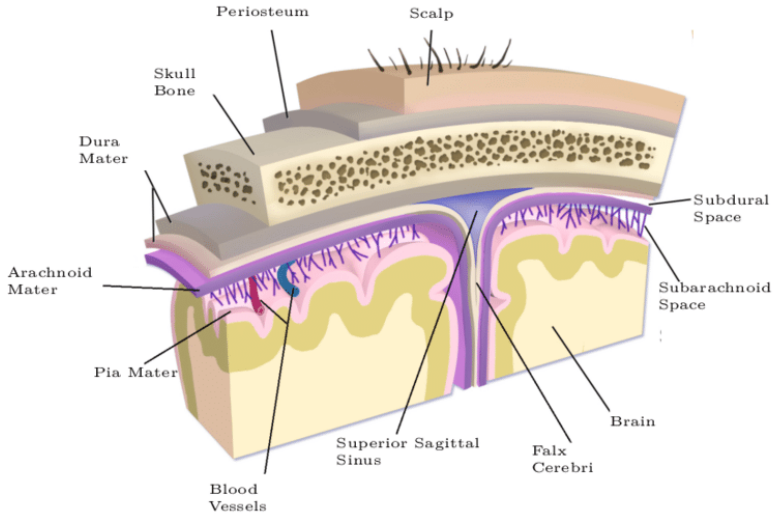


Figure 2.5: Illustration of the layers in the brain and head (Image source: [30])

Table (2.1) shows the thickness and resistance of the different layers of the head that the electrical signals are measured through. The layers has the properties that is the reason why EEG signals are generally nonlinear as well as non-stationary. The signal may change over time in terms of mean, variance and higher order, but it is possible to quantify the signals by measuring some statistics of the signals at different time lags. It is possible to observe these changes during eye blinking, the transition between alertness and wakefulness, in the Event-Related Potential (ERP) signals as well as during the transitions between various middle stages of a seizure. ERP is a technique used similar to EEG, where the difference is that in ERP the researcher is looking for activity related to a specific stimulus the participant is presented for [5].

2.6 EEG recording and measurement

When it comes to recording brain activity there are three methods that are more common and frequently used, where the motivation for using EEG is that with this technique the electrodes are placed with a non-invasive procedure, with limited additional equipment needed. This is in comparison to one of the other methods called Electrocorticography (ECoG), where the electrodes as attached directly on the surface of the brain during a medical invasive procedure. The number of channels used in EEG can range from only a single channel to as many as 256 channels, where a cap is common for a larger number of channels to make the placement of the channels easier and less time-consuming. There will be the need to make some adjustment with a cap based on the size of the head, and the amount of hair the participant has. But nevertheless, the placement will be fairly equally placed. If a higher disturbance caused by hair or the electrodes are not properly

Name of band	Frequency limits [Hz]	Associated with
Delta (δ)	0 - 3.99	Deep sleep
Theta (θ)	4 - 7.99	Drowsiness
Alpha (α)	8 - 12.99	Relaxed awareness
Beta (β)	13 - 35	Active thinking
Gamma (γ)	> 35	Attention and memory

Table 2.2: EEG frequencies in regards to scoring, defined by AASM

attached, the channels affected can be replaced by near-by channel and data will not be lost. There are also electrodes attached to the face to monitor head- and eye-movement.

The limitations for using EEG is that the signal is vulnerable to noise which may be caused by participants moving their head during the recording, creating space between the electrode and the scalp and therefore disturbance, or eye-blinking which has a very distinct shape when looking at the recordings and may therefore disturb what the researcher is looking for. The electrodes are also sensitive to other electronic devices within proximity, and there most therefore be taken precautions. The spatial resolution is relatively low, which makes it more difficult to find the location to the source of the signal.

The last common method is Magnetoencephalography (MEG), which is also a non-invasive technique. However the MEG requires a relatively large machinery, and therefore makes EEG the preferred choice when it comes to brain recordings.

2.7 Frequency bands of the brain

When analyzing brain activity with EEG, the recordings are divided into five frequency bands, given by table (2.2). Visualization of the different bands can be seen in fig. (2.6). The bands are typically associated with specific brain states, however due to variations of definitions there is not possible to translate them into specific brain processes [20]. Age is a factor that will change the characteristics of the waves, while amplitudes and frequencies of the brain rhythms may vary from one state of a human to another [55].

Artifacts are different responses that may interfere with the signal and give impression that there are other responses happening due to muscle movements or electrical devices nearby. These can be removed by applying different signal analysis, which will be mentioned in section 2.11.

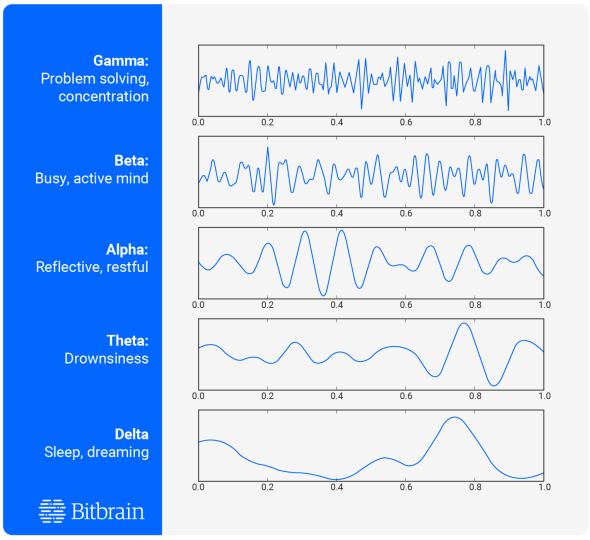


Figure 2.6: Illustration of different frequency bands (Image source: [7])

2.8 Sleep classification

Scoring a participants sleep cycle is normally done by a sleep expert through visual inspection of the PSG signals, and thereby scoring them based on a fixed 30-second epoch according to the AASM manual. The fixed length of 30-seconds goes back to 1938 where Loomis et al [3] used a "paper cutting brain potential recorder", which is a roll of paper that was cut by a knife each 30 seconds. The Rechtschaffen & Kales (R&K) manual also accept epochs with the fixed length of 20 seconds, while the AASM only recommends 30 seconds.

When looking at the PSG the sleep expert will look for specific characteristics of the different sleep stages and annotated according to this. If an epoch has more than one stage-characteristic present, the most prominent stage in terms of time will be the one annotated. The scoring and classification of a participants sleep cycle is a very time-consuming work, as well as monotonous work. An eight-hour recording of a sleep cycle may require two hours of work for the sleep expert [67], where further classification of events will add on this time spent. For a sleep lab they will need data from multiple participants, where the same repetitive work will be needed for all of them before the actual analysis of the recording can begin.

To check the reliability between the two different scoring definitions, R&K and AASM, a study in 2009 concluded that there was an increase in overall agreement between the methods from 80.6% and 82.0% [14], two years after AASM updated their classification rules. The biggest disagreement between them was in terms of annotation related to stage N2. In 2013 "The American Academy of Sleep Medicine inter-scorer reliability program: sleep stage scoring" presented their result where 2.500 sleep scorers had participated [51]. Most of the scorers had three or more years of experience. The overall agreement for sleep stages were 82.6%, with the highest agreement was for REM sleep, followed by N2 and WK at the same level. The lowest results were for stage N1 with 63.0% agreement among the scorers, with stage N3 at 67.4% agreement. This shows that the updated version of AASM is still not sufficient enough to ensure accurate scoring.

2.9 Problems with today's sleep classification

The practice of today is based on definitions from two different manuals, as well as the scorers subjective perception about what stage the different epochs are representing. As already mentioned, this practice is time-consuming, imprecise and repetitive, and in dire need for technological advancement. Sleep scorers might also set different start times and end times for the annotated spindles, which will lead to a large variability in spindle annotation. The interpretation of characteristics and definitions may lead to different staging on same datasets completed by different scorers as seen when 2.500 scorers looked at the same data [51]. The overall agreement result is a satisfying number, but there were bigger dispute about some of the sleep stage as well as specific transitions between some of the stages which tells that there is need for better clarifications. An alternative to increase the accuracy is to have more scorers look at the same date, and assess the data together to end up with a result supported by more experts. This may increase the accuracy of the scoring, but will also require more manpower and may use more time. However, the human error will never be removed, and is a component exposed for numerous possible distractions. The only way to remove the human error is the remove the human entirely from the sleep classification and rather place them at the verification stage for the detectors.

Utilizing the enormous amount of automatic detectors and classifiers that has blossomed the past years are a way of removing the tedious work, and focus more on the results and how to employ this. The detectors found public will perform at the same level as an sleep expert or with a higher accuracy [10], depending on datasets, models and focus areas. Using the detectors are also a great tool to train sleep experts by comparing the two sleep classifications, and from this continuing developing better working models and experts. Many of the automatic sleep scorers are available online, but some of them are largely dependent on having a bigger dataset available. Due to restrictions in terms of privacy and confidentiality, there are not many public dataset available for researcher which therefore often end up using the same ones. These are often sampled at a

low frequency, as well as only contain a smaller amount of channels. This limits the possibility to explore other events in the dataset besides staging, as well as alternating between the channels to gain the best result if one wants to find the origin of a signal. Many of the models often have a small dataset following their model so a researcher can run and validate their result, but not necessarily add on their own data without the need to adapt this to the model.

Another challenge is that most of the public data available are gathered from healthy adults at a certain age, and therefore the easiest basis to train different models with. However will characteristics such as age, gender and sleep or neurological disorders affect the PSG recordings, where these recordings also might be different from one day to another for the same participants. Models trained with this will therefore have a low transfer capability, and studies show that models trained on healthy participants will predict results with a drastically lower accuracy for participants with for instance sleep disorders [12].

The gold standard of today is PSG records which consist of mainly EEG, EOG and EMG to evaluate the sleep cycle [52], with visual inspection to define stage and events. However is this impractical for long-term studies as well as home utilization since it often requires laboratory equipment as well as being vulnerable to noise from for instance other electrical appliances [62]. For a participant taking part in an sleep study, the first night is often excluded since the PSG recordings often show a decrease of total sleep time as well as lower sleep efficiency, in addition to other factors. Normally the unrepresentative sleep pattern is not a part of the study, where the participant is asked to return for the actual recording. Still may the first recording be scored to make sure the participant meet the requirements to be a part of the specific experiment.

2.10 Sleep spindle

AASM defines a spindle to be in the frequency range of 11 to 16Hz, where 12 to 14Hz is the most common frequency to focus on when looking for spindles. The duration for each spindle is around 0.5 sec and up until 3 sec. Sleep spindles are most prominent in the N2 sleep stage, and is one of the characterising when scoring sleep for the same stage [49]. The spindles play an important role of sleep in memory and learning, where the propagation into the hippocampus reactivate freshly encoded reflections. This will again enable effective transmission and storage of reactivated memory information within the neocortical network [60].

Spindles may also be divided into two subcategories - *slow* and *fast spindle*. Fast spindle is defined to have a frequency between 12 to 15Hz, where they for healthy young people often are synchronized with the depolarizing Slow Oscillation (SO) (<1Hz) up-state [46]. Slow spindles are categorized to be between 9 to 12Hz, and often occur at the transition into the SO down-state. An illustration of how spindles may appear in correlation with SO is added in fig. (2.7).

Sleep spindles are shown to appear between 4 and 6 weeks of age [56], where the density of spindles decline as one gets older and the sleep cycle changes. The change between the spindle density amongst young and older men are more apparent then compared to the decline for females [47]. As a person gets older, the deep sleep will be more difficult to reach, and will therefore impact how frequently the spindles appear as well as their appearance.

2.11 Artifact removal

Artifacts are undesired signals that may contaminate the signal of interest by introducing changes in the measurements [29]. Best practice when working with EEG is is to avoid occurrence of artifacts, but unfortunately various physiological factors may interfere. Some of the most common artifacts are cardiac activity, ocular movements, eye blink and muscular activity. The cancellation of noise and artifacts are important issues in EEG signal processing. Removing these artifacts may

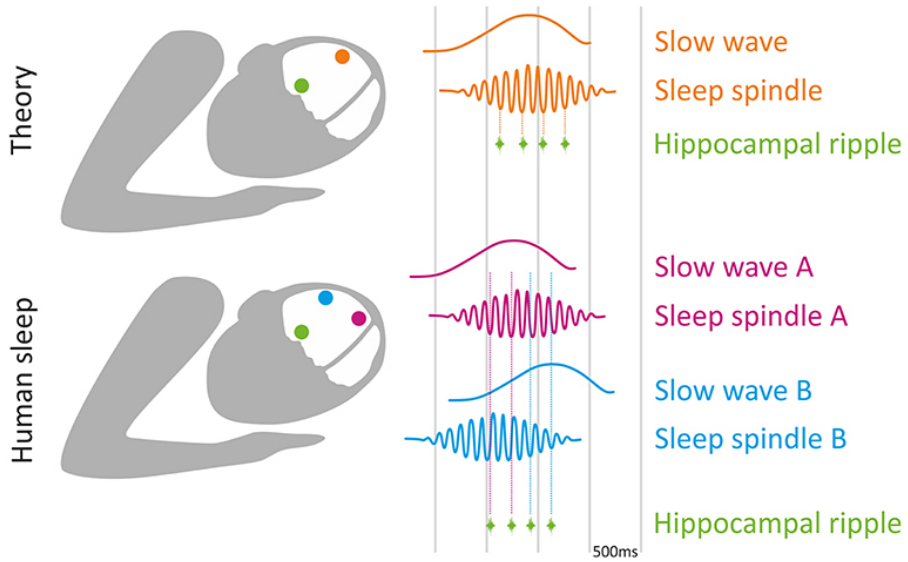


Figure 2.7: Illustration of how spindles may occur in correspondence with SO (Image source: [60])

be to create markers to identify segments with poor quality, or to cancel artifacts from the EEG signal. Rejection of segments are not desired since one wants to retain as much data as possible.

There are many different methods and techniques to remove artifacts, but there is however not an agreement among data researchers on an optimal method to improve the quality of the recorded EEG signal [29]. This is mainly due to the many different characteristics that can be found in the signals, as well as the scarce available public datasets.

2.12 A7

The detection model called A7 (algorithm #7) was developed as a simple and efficient sleep-spindle detector, with the motivation to emulate human scoring [35]. The model runs on a single EEG channel, where four parameters needs to exceed set thresholds simultaneously. In the study, comparing the scoring of the model to an individual sleep expert the model performed just as good as well as surpassing other spindle detectors. The detector focuses on minimizing the number of "hidden spindles", which many models tend to find. These hidden spindles are false-positive labels, which can not be seen in the raw EEG signal and are found in all stages.

2.13 DOSED

In 2019 a novel deep learning architecture called Dreem One Shot Event Detector (DOSED) was introduced, specialized in detecting sleep related micro-architecture events [10]. The approached was initially tested for three types of events - *spindles*, *K-complexes* and *arousal*. The results based on four datasets were then compared to the current state-of-the-art detection algorithms, where the versatile method outperformed the latter. Earlier detection models for these events have been based on band-pass filtering of the EEG signals, as well as using hand-crafted features. The current model is inspired by object detectors developed for computer vision such as YOLO [9] and SSD [54], where the convolutional neural network builds a feature representation from the raw EEG signal together with two modules performing localization and classification for the events.

Chapter 3

Methods

> Human-in-the-loop *(HITL)* evaluations are imperative to delivering accurate, relevant information and avoiding bias. Despite what many believe about humans actually taking a backseat in AI training, I think we'll see a trend towards HITL evaluations in an effort to empower responsible AI [16]
>
> *Sujatha Sagiraju*
> *Chief Product Officer, Appen*

The proposal for this paper is to exploit two different software detectors to be able to create a better individual database, where the use of sleep scorer is set to a minimum with verification as the main job instead of spindle annotation. The motivation is to be less dependent on the sleep expert for processes where automatic detectors can be used as support or work as a replacement. To begin with will this process require some time, as well as comparison between the different models with an experts opinion and annotations. However will one with time be able to limit the use for the sleep scorer, where they can use their time on other matters beside annotation. This is obtained by utilizing the Lacourse Spindle Detector (A7) model to build up a database with automated spindle annotations similar to a human scorer from preprocessed data, and then using the DOSED to locate spindles for unknown raw PSG recordings. The sleep scorer will then use their time for validation parts of the PSG recording, which is far less time-consuming. DOSED will run by itself and requires minimal interference to be able to run.

3.1 Artifact removal

3.1.1 Filtering and resampling data

The IIIS dataset was resampled and filtered before further analysis was completed to limit unnecessary memory allocation, as well as removing artifacts at specific frequencies. Downsampling is one of the options with resampling, and is performed with a low-pass filter applied to prevent aliasing followed by decimation where every N^{th} sample of the signal that is selected [41]. Keep in mind that edge artifacts will be introduced, which is a problem when working with epochs and this is then applied to each start and end. Therefore it is recommended to downsample the raw data before dividing it into epochs, which will only have edge artifacts at the start and end of the entire recording. Downsampling is often done for signals where time precision is not crucial as well

to achieve a faster computation where the experiment it self is sampled with a high frequency. For this project the raw data was downsampled to 256Hz, from the original 1024Hz [44].

Following the downsampling, filtering was applied to remove possible affects caused by power-lines [36]. The power-line noise is created by electrical network, with a sharp peak at either 50Hz or 60Hz [41]. The frequency is depending on the geographical location. There may also be some peaks present at the harmonic frequency such as the integer multiples of the power-line frequency. For frequencies of 50Hz, the harmonic frequencies would be 100Hz, 150Hz and so on. For an initial power-line peak at 60Hz, the harmonic frequency will then be 120Hz, 180Hz and so forth. To remove this "Notch filter" was applied, a bandstop filter with a narrow stopband which is also the most common way to filter data at respective frequencies [43]. Be aware that filtering may cause distortions in the pass band as well in the resulting time domain signal, which may produce artifacts like ringing. These artifacts can be seen as "ripples" in the time domain [64]. The "Notch filter" was set to 50Hz, before a bandpass filter also was applied with the frequencies set to 0.3 to 50Hz.

3.1.2 Independent Component Analysis

As mentioned, there are several different ways to remove artifacts from the raw EEG signal, where Independent Component Analysis (ICA) was selected to separate independent sources linearly mixed without removing the affected data portion [37] [17]. ICA is a part of the open-source package called "mne", which is used to analyze and visualize neurophysiological data [21]. The electrodes attached to the face and scalp will for instance record blinks, heartbeats, activity in different areas of the brain as well as jaw clenching and swallowing [45]. As lang as the various signals are statistically independent and non-gaussion, ICA will be able to separate the sources and reconstruct the signal after the unwanted signals are excluded. ICA was initally made to work well on multi-channel EEG, but studies show that it is also a good option for cases with only a few channels [50].

In short an ICA object is created, before the object is fit to the data. From there the unwanted components are excluded, before the ICA-fit is applied to the raw data or epoch. An illustration explaining how ICA-fit is computed can be seen in figure (3.1).

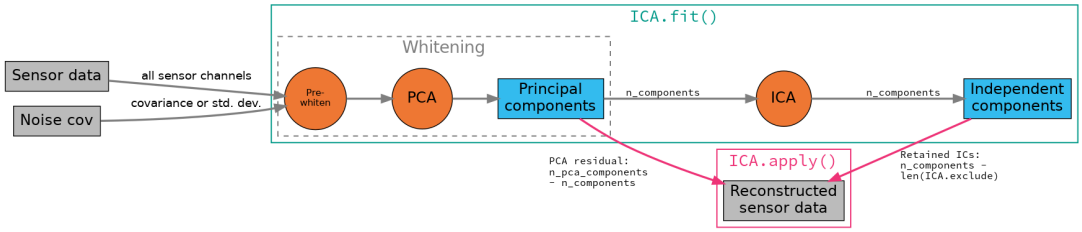
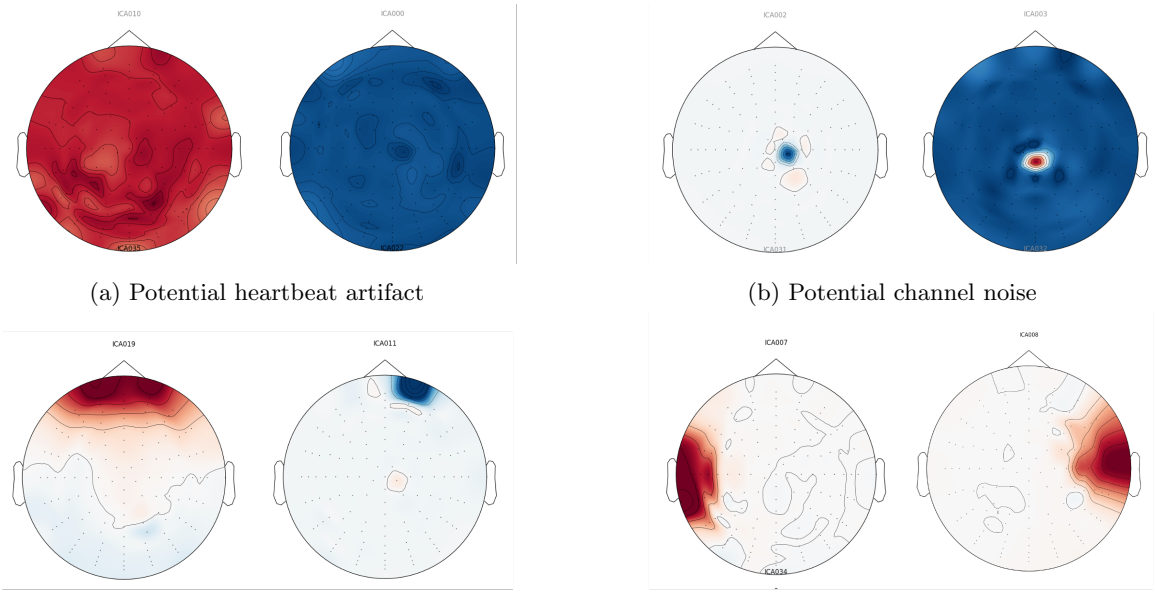


Figure 3.1: The process of ICA.fit() [Image source: [37]]

In detail the ICA object is created with a defined number of Principal Component (PC)s, where the number is selected based on the explained variance to be higher than 95%. To be able to reach this the number of PCs was set to the range of 40-50, depending on the resulting explained variance. The components were then visualized by the use of `ica.plot_sources(filtered_data, show_scrollbars=True)` and `ica.plot_components()`. Both of these were used to confirm the exclusion, and therefore remove those who were the most prominent rather than remove to much and therefore possible valuable information. This process is based on visual confirmation, looking for specific characteristics that stands out form the rest. This could for instance be a heartbeat visible in the sources, with a repetitive spike throughout the recording, or other abnormalities. Looking at the generated head-models for the different components, there are some characteristics to look for that were present for all recordings of the participants with the most prominent ones indicated by fig. (3.2) [34].

The electrical potential generated by the heart is captured as a heart component. Examples of



(a) Potential heartbeat artifact

(b) Potential channel noise

(c) Potential eye movement

(d) Potential muscle artifact

Figure 3.2: Potential artifacts that should be removed when using ICA

heart artifacts found when cleaning the IIIS dataset can be found in fig. (3.2a). The signal is called Electrocardiogram (ECG), and is sometimes referred to as a pulse artifact. The amplitude is low on the scalp, but may cause rhythmic distortion on the EEG signals and therefore removed.

Channel noise may occur if a channel gets bumped during the recording, or if there is a poor contact between the electrode and the scalp. This may cause larger artifacts only located in this specific channel. ICA will do this separation and create an own component. An illustration of this can be seen in fig. (3.2b).

Eye-movement or blinking may be seen in fig. (3.2c), where the retina creates an electric field that is captured by the electrodes. Eye-movement can be split into two components - *vertical movement* and *horizontal movement*. The model to the left indicates the effect eye blinks has to the signal, with equal colors and intensity on each side of the nose. The right model shows an eye movement due to its origin close to the one eye.

Fig. (3.2d) shows indications of muscle movement. The concentration of the scalp is darker closer to the ears due to the EMG channels placed at the earlobe. This may cause high frequency signals that overlap with the EEG signal, and is therefore removed during the cleaning.

Since EOG channels were a part of the PSG recordings, eog_indices, eog_scores = ica.find_bads_eog(filtered_data), was also used to determine which components to exclude [42]. This process is also a more convenient way to make automatic marks for exclusion. While the previous visual components are based on manually confirmation of characteristics, and therefore not as easy to exclude automatically. If there is no EOG, one can use ECG in almost the same way. There is also possible to use a single channel, or create a bipolar reference from the frontal EEG sensor as a virtual EOG channel. However this may cause some problems since the selected EEG channel only can reflect EOG and not brain dynamics in the prefrontal cortex. Fig. (3.3) shows the components that are marked as bad in red bars and therefore should be removed.

When the excluded components were defined, the ICA was applied to the raw data, and a new run was initiated to see if there were other components that should be removed. This work is time-consuming due to the processing and fitting of the ICA requires memory to be able to run, as well as being a subjective selection done by the researcher which is what we want to exclude from the total model. There are other methods to remove artifacts more automatically, but until there is a bigger consensus among researcher and more models to utilize, this method is the best for this purpose. Nevertheless, using time getting to know the data, and use this as a learning

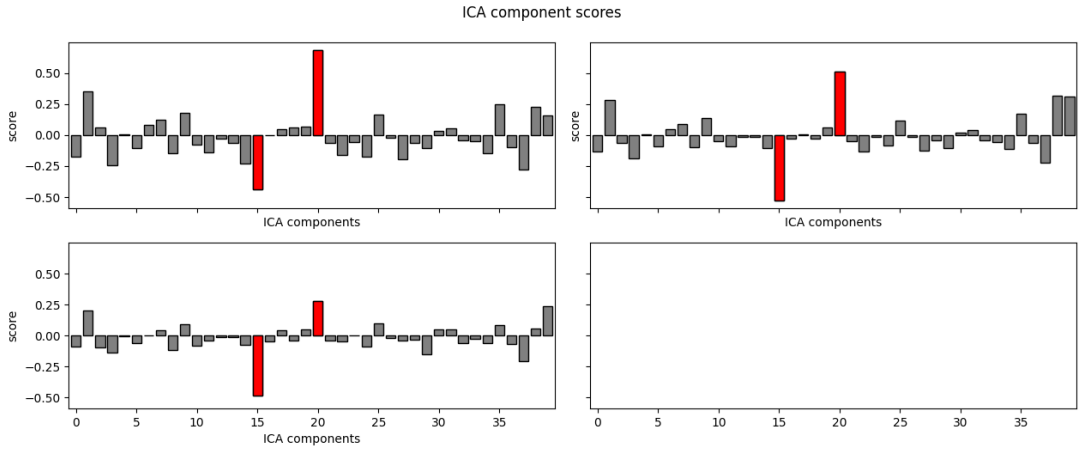


Figure 3.3: The results for one of the subjects while removing artifacts

experience for new researchers entering the field will also give a lot of insight into how to perform later experiments. Specifically in terms of being aware of noise in the surroundings, as well as how muscle movements and heartbeats interfere when focusing on brain activity. The attention to such details, may lead to creating also a better atmosphere for the participant so they can relax their muscles.

3.2 A7

Pros	Cons
No annotations	Preprocessing required
Fast computing	Artifact-free data
Single channel	Sleep staging required
Mathematical-based model	Three input array

Table 3.1: Pros and cons table for using A7

3.2.1 Model

A7 is a mathematical-based model, looking for events where four thresholds are exceeded simultaneously indicating the sleep spindle [35]. The model was chosen since it did not require spindle annotation, and therefore independent when comparing how a sleep expert would score the same dataset. However did the model require clean data, meaning that all artifacts needed to be removed to make sure this would not compromise the calculations. The cleaning of the dataset is described in section 3.1. Nevertheless, the preprocessing is time-consuming and since it was done manually it would be imprecise due to biased opinion. With the current development in sleep research, it should be possible to automate this process and therefore remove the human error.

The model is derived from similar detection algorithms, and was tested against a gold standard spindle dataset based on a consensus of a group of human experts. The results for the detector was 75% precision, 68% recall and with an F1-score of 0.70. Comparing this to an individual human expert with an average F1-score of 0.67, the model is relatively equivalent and therefore a good use for our model.

The model extracts four parameters to be able to detect spindles, based on the broadband filtered EEG signals (0.3-30Hz) and the sigma band pass filter (11-16Hz). For the purpose to detect both fast and slow spindles, the filter was set to 9-16Hz. The four parameters set as threshold for the model are *Absolute sigma power*, *Relative sigma power*, *Sigma covariance* and *Sigma correlation*.

The model is set to detect a spindle if the parameters above is exceeded at the same time, as well as following the guidelines from the AASM manual regarding spindle detection. However, the initial model is set to find spindles even if they are as short as 0.3 seconds, compared to the recommended duration from 0.5 seconds. This was done due to the reliability that experts may be able to see spindles as short as the first mentioned, as well as having the same characteristics in terms of oscillation frequency, amplitude and symmetry found in the study from 2014 [63]. At the same time are short events merged together if the individual events are shorter than 0.3 sec and less than 0.1 sec from each other. It was still required that the merged event would be a minimum of 0.3 sec and maximum of 2.5 seconds to be considered a sleep spindle. For a human sleep expert, detecting these spindles and annotating them would be prone to bias relative to the closeness of the events. In this report the guidelines of AASM were followed in terms of duration of spindles, and therefor set from 0.5 sec to 3.0 sec.

The user of the model needs to provide three arrays containing data from a single EEG channel, artifact detection as well as sleep stages. The data from the channel needs to be the unit μV, while the artifact array is a list containing 0 or 1 based on whether 30 seconds of data around this point should be excluded from detection. More details about how this was utilized can be found in subsection 3.2.2, as well as sleep stage array.

Absolute sigma power

The absolute sigma power *(A7absSigPow)* is used to identify train of sigma waves, and is computed through a Power Spectral Analysis (PSA). N defines the number of samples in the window length, set to be 0.3 seconds initially to be able to detect the shortest spindles. The unit of the parameter is $\log_{10}(\mu V^2)$, where i indicate the index sample of the EEG signal band passed filtered in the sigma band. The equation for the absolute sigma power can be seen in equation (3.1).

$$A7absSigPow = log_{10} \sum_{i=1}^{N} \frac{EEG\sigma_i^2}{N} \tag{3.1}$$

Relative sigma power

The relative sigma power *(A7relSigPow)* is also computed using PSA, and is defined as the ratio of the power in the sigma band over the broadband excluding the delta band (0.3-4.5Hz). The equation for the relative sigma power can be seen in equation (3.2), where the baseline only consist of artifact-free data, and may therefore be restricted to specific sleep stages.

$$A7relSigPow = zscore\left(log_{10}\left(\frac{PSA_{11-16Hz}}{PSA_{4.5-30Hz}}\right)\right) \tag{3.2}$$

Sigma covariance

The sigma covariance *(A7sigmaCov)* parameter is based on the covariance between the broadband signal *(EEGbf)* and the signal filtered in sigma *(EEGσ)*, where these two inputs need to have the same number of samples which is defined by the window length. This is since the samples in EEGbf is matched with the samples in EEGσ, and the covariance is the average product of a paired sample, as shown in equation (3.3). A high covariance is obtained when EEGbf and EEGσ are varying with the same polarity and can varify from -∞ to ∞. The sigma covariance is also based on data without EEG artifacts, and therefore may be limited to specific sleep stages. The equation is shown in equation (3.4), where the sigma covariance is z-score of the log 10 transformation of the covariance between EEGbf and EEGσ.

$$cov(EEGbf, EEG\sigma) = \frac{1}{N}\sum_{i=1}^{N}(EEGbf_i - \mu_{EEGbf}) * (EEG\sigma_i - \mu_{EEG\sigma}) \tag{3.3}$$

$$A7sigmaCov = zscore(log_{10}(cov(EEGbf, EEG\sigma))) \tag{3.4}$$

Sigma correlation

The last parameter is the sigma correlation *(A7sigmaCor)*, defined by the covariance normalized by the amplitude of both variables. The values of the equation (3.5) is in the range from -1 to 1, where it typically lays between 0 to 0.5. This will result in a normal distribution of values, and therefore not log transformed. Standard Devitation (SD) refers to the standard deviation of the two variables.

$$A7sigmaCor = \frac{cov(EEGbf, EEG\sigma)}{SD_{EEGbf} * SD_{EEG\sigma}} \tag{3.5}$$

3.2.2 Features

This model focuses on finding sleep spindles to emulate human experts, and is not applicable to other events besides spindles. The thresholds selected to be exceeded to indicate spindles for this project were the same as the ones provided in the paper [35]. These parameters had been tested on several datasets, and designed to detect spindles that are clearly distinguishable from the background.

Initially the model was used to detect spindles in N2 sleep, but was also utilized for detecting spindles in N1 and N3. REM and WK were excluded by using the input array requesting artifact information. Corresponding with the sleep stage scoring of the participant, an artifact array was created were each epoch was set to be 0 or 1. The value 0 corresponded with NREM *(N1, N2 and N3)*, while 1 corresponded with stage REM and WK meaning that these would be excluded when running the detection model. This was a way to ensure that the data was kept whole in terms of time domain, and therefore easier to work further with. WK and REM was excluded since they will contain frequency bands that will be similar to a sleep spindle, but spindles are defined to be in NREM [18].

The sleep stage used as input was set to N2 for the entire dataset, instead of using the initial scoring provided by either an expert or automatic sleep scoring detectors. This was done to match the default parameters in the code, as well as to be able to detect over several sleep stages. Another motivation to do it this was to be less time-consuming, where an option could be to separate NREM from WK and REM in this array instead of using the artifact array.

3.2.3 Preprocessing

The model required artifact-free data, where the best results would be yielded from clean spindles with high signal-to-noise ratio in the data.

The cleaning of the IIIS data was completed for the eight subjects, where five of the subjects went through one or two rounds with ICA, where approximately six components where removed the first round and three to six where removed the second round. The cleaned data was then run through the A7 detector to verify the result together with a sleep expert.

The remaining subjects went through several rounds of cleaning, where each round was tested with the A7 algorithm to detect spindles. Without satisfying numbers, the data would be sent through more rounds of cleaning. The same procedure was applied to all datasets, focusing on the same characteristics.

3.2.4 Channel selection

The A7 algorithm is based on running from a single EEG channel, where channel C4 was used for this project. This was done since this is often the channel used for detection by sleep experts, and therefore would provide comparable results with already annotated spindles. For the IIIS datatset 128 channels were available for selection.

3.3 DOSED

Pros	Cons
Fast computing	H5-file
Multiple running subjects	Annotations needed
Easy to use	Dataset with same amount of channels
Machine learning based model	Equal channel names

Table 3.2: Pros and cons table for using DOSED

3.3.1 Model

DOSED is a machine learning based model, with the possibility to run on several unknown dataset to detect spindles or other events [10]. This model was chosen for the project since it could run on several unknown datasets, and therefore would save time for detecting spindles. There is also no need to clean the data, where the user only need to provide H5-files containing the spindle annotations as well as the data from a defined number of channels. For the initial code, all datasets need to have the same number of channels, as well as they need to have the exact same names to be able to run.

By using the localization module, it is possible to predict duration and centers for potential events while the classification module predicts the corresponding label. The architecture is trained end-to-end by back-propagation. The module looks at a short window of signal extracted from the PSG recording, where the window time is approximately 30 seconds. The model divides the segment into default events, where one event lasts for about 1 sec, and is set every 0.5 sec and therefore has some overlapping. The overlapping between each default event is chosen since this is a typical duration of a event, where this method focuses on events <0.5 sec. An illustration of this method can be found in fig (3.4).

Depending on the events one is trying to localize, there might be changes based on duration and overlap. The model predicts potential events associated with each default event, including the probability of the event having a label. The potential event with the highest probability label, as well as when the probability is higher than a specific (cross-validated) threshold, are selected. Thereafter is a non-maximum suppression applied to remove overlapping events, where the motivation is to group predictions with the same label based on their Intersection over Union (IoU) and keep the highest probability of observing a true event in each group.

To be able to run on a computer that does not have Graphics Processing Units (GPUs), the code for creating the network need to be changed to be able to run on Central Processing Unit (CPU). This was done by changing the device to `device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')`.

The downside of this model is that most of the default events may not match any true events. The algorithm wants to maximize the number of true events, which means that for a specific EEG sample, there might be at most one third of events considered to be negative defaults and therefore does not match a true event. However to cope with imbalance in the dataset should at least 10 default events be used as negatives, and to improve the training of the approach compared to random sampling strategy.

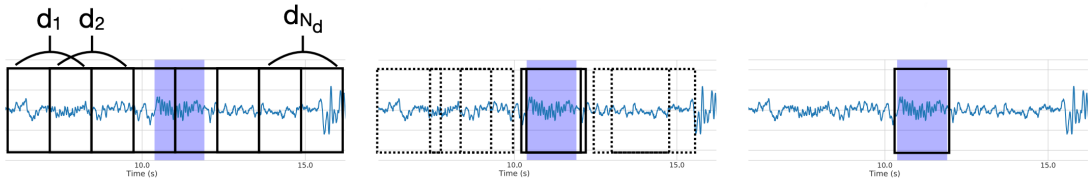


Figure 3.4: The prediction procedure used for DOSED [Image source: [10]]

3.3.2 Features

The motivation to use this detector is the ability to look for multiple events, as long as data contain the location for the events in terms of start and end time. For this project, spindle detection was the only event focused on, but the model has also been tested for k-complexes as well as arousal [10].

3.3.3 Preprocessing

There is no need to preprocess any data, since any processing needed is already a part of the code. To deal with the computation time and to use more channels, the IIIS dataset used for this project was resampled from 1024Hz to 256Hz. Since the method was based on using two channels with a sampling frequency of 64Hz, was this also tried to see whether this would have any impact on the results.

3.3.4 Channel selection

Due to the limitation of data two different approaches were tested regarding channel selection. The code was initially developed to use two channels, where C4 and B22 were chosen from the dataset due to their placement and use in detection and scoring. In addition to this all 128 EEG channels were also tested. One needs to keep in mind that all data needs to have the same number of channels, where the current model does not open to alterations on this. All channels also need the same labeling for all data used, where this needs to be altered before creating the H5-file.

Chapter 4

Results

The motivation for this project was to find an alternative method to detect and predict spindles, with the interest of exploiting software detectors instead of annotations made by expert. More information about the dataset is presented in terms of distribution of sleep stage as well as visual representation of each subjects sleep cycle. Further is the results from the software detectors, as well as how the impartial platform can be more cost effective in the future.

4.1 Semiautomatic spindle detection platform

The idea behind the semiautomatic spindle detection platform with two software detectors arose when testing different spindle detectors with the data from IIIS to obtain more data with spindle annotation. The methods found online either required a large dataset to train on (DOSED), where annotations were a requirement, or were time-consuming and required a lot of preparation to be able to run (A7). Joining the IIIS data with annotated data available online did not get satisfying answers due to them being artifact-free in contrast to the data used in this report. From this the idea of building up a database with the artifact data began, where the disadvantages for DOSED about annotated data could be obtained by running A7 in advance in a form of pipeline method. In addition did this method not require much work from a sleep expert, and were designed with the conclusion that the sleep expert would work with verification instead of a annotation.

Figure (4.1) shows the steps of how the initial process emerged, with the basis of having comparable annotations made in this case from a sleep expert. These annotations were compared with the spindle detections from A7, were further tuning and changes to A7 were based on the correlation. When the detections were satisfying enough, the annotations were united with the raw data resampled to either 64Hz or 256Hz. Unknown data were used to test the results from DOSED, where the sleep expert only would receive 10% of the data to verify. Feedback from the sleep expert lead to changes for parameters in the machine learning method and a final result for the unknown data could be presented. Keep in mind that DOSED requires diverse database, and will not be able to give accurate and satisfying data without a good assortment of subjects. The data used in this report were from all females within the same age group.

The next chart, fig. (4.2), shows the process when working on data without previous, comparable annotations. This was relevant when working on the six subjects from IIIS that did not have previous spindle annotation, and most likely for new data presented. The changes between the charts is the comparison, where tuning of A7 were based on the previous results with comparable annotations. The tuning is referring to changing the threshold parameters. In this chart the sleep expert would only have a role of verification after running DOSED. This implies that the preprocessing of data is successful and that the A7-algorithm gives expected number of spindles which are based on previous annotations as well as reference to other papers regarding what to expect for spindle detection.

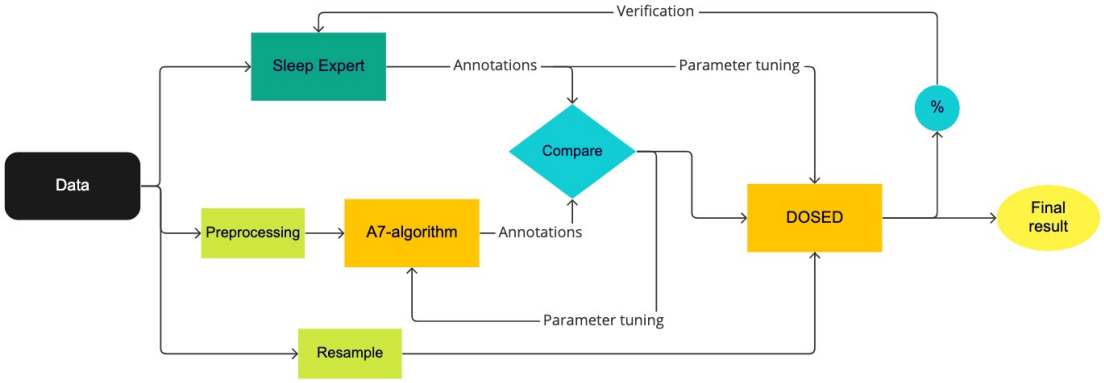


Figure 4.1: Flowchart representing how previously annotated data were used in the proposed model

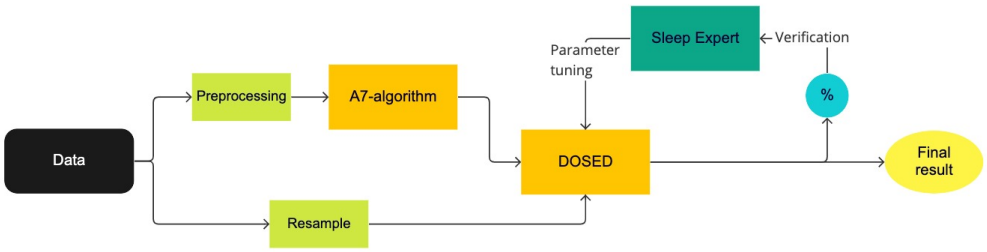


Figure 4.2: Flowchart representing data without previous annotations

The final project with time would only be to run DOSED, where predicted annotations from unknown data would be added to the dataset. There would therefore not be any need to continue run A7. However could this be used to test random samples of data to eliminate possible bias in the database since it is independent from previous calculations. This would then be run simultaneously to DOSED and compare the output to make necessary changes. The method may not be able to predict all spindles in the data, but should give some of the most prominent ones.

4.2 System specification

For most of the computation done in this project was completed with a MacBook Air, with a 1.6 GHz dual-core Intel core i5 processor with 16 GB memory and 500 GB storage. Some computations were run using Idun cluster, a project between NTNU's facilities and the IT division [38]. This cluster was mainly used for resampling, in addition to computers with a memory around 64 GB. This was done due to the limitations with memory, while most of the prediction could be done on the specified computer.

Due to the limitations for computers memory, the computation time will vary for different computer and should be taken into account. However are most of the calculations and predictions done with these specifications, which also shows that much can be done without requiring large amounts of equipment and computer power. That makes this model versatile and convenient, but will need more storage as the database grows.

4.3 Performance

4.3.1 Sleep stages

The distribution of sleep stages can be seen in fig. (4.3), for all eight subjects in the IIIS dataset. In overall the subjects seems to have an adequate sleep pattern, with a few exceptions. "Sub 0129" has the most common distribution which is 5% in stage N1, 50% in N2 and 20% in N3 sleep [57]. However has "Sub 0242" a distribution that is almost alike for all stages. This uniformly distribution is a bit unusual with only 9.7% difference between the highest occurrence (N2) and lowest (WK). "Sub 0241" is the subject with the longest time in N2 sleep, as well as the longest time in N1 sleep of all subjects a part in this project. "Sub 0248" has the most epochs with a total of 1011, where "Sub 0247" and "Sub 0300" are equal at 987 epochs, which is the lowest number.

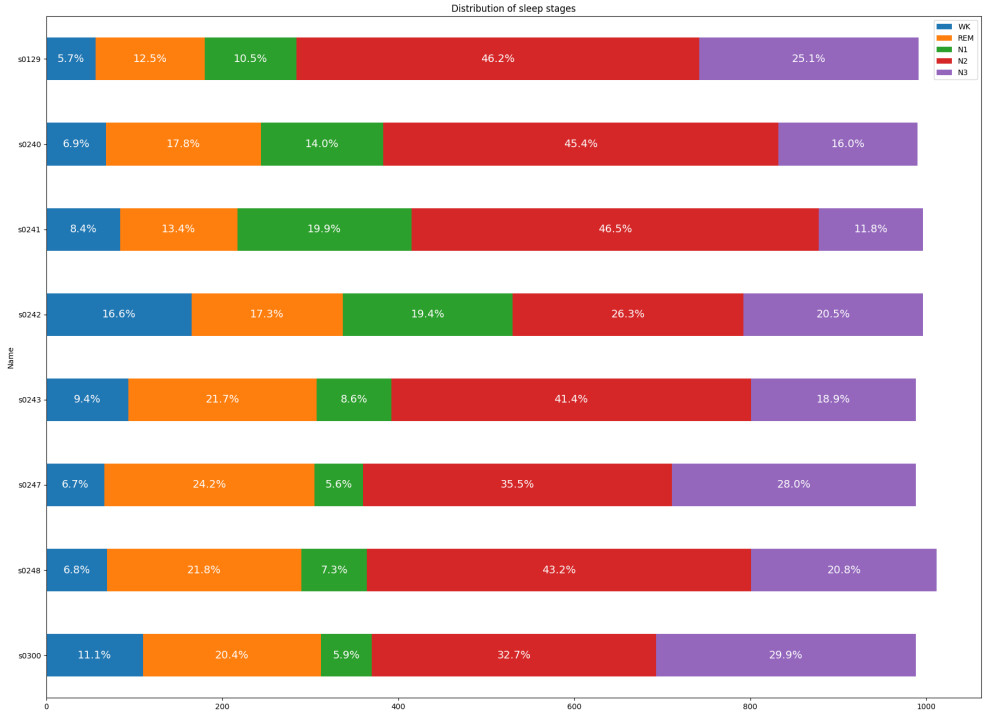


Figure 4.3: Distribution of sleep stages for all subjects

Figure (4.4) shows the sleep cycle for all subjects, where "Sub 0242" stands out due to not reaching N3 sleep within one hour (3600 sec) as the other subjects. However is the subject alternating between WK and N1 sleep, with three times reaching N2 before it re-enters lighter sleep and awake state. The subject needed approximately 1.5 hour (5400 sec) from the recording began to it reached deep sleep at N3. "Sub 0248" has the least annotations for WK during the sleep cycle, with a higher occurrence closer to the end of the time series. WK stage is indicated by the blue color, while REM is the orange color. Stage N1, N2 and N3 are colored subsequently in green, red and purple.

4.3.2 A7

The data collected at the Human Sleep Lab at IIIS consisted of raw PSG recordings, where the data needed to be cleaned before running the A7 algorithm. From the total of eight subjects, five yielded good results, while three of the subjects ended up with very poor number of spindles or none spindles. This can be seen from the third column in table (4.3). Only "Sub s0129" and "Sub 0240" were annotated by the sleep expert, and the remaining six subjects only contained a sleep stage scoring. The same procedure for removing artifacts were applied for all subjects, following the same procedure.

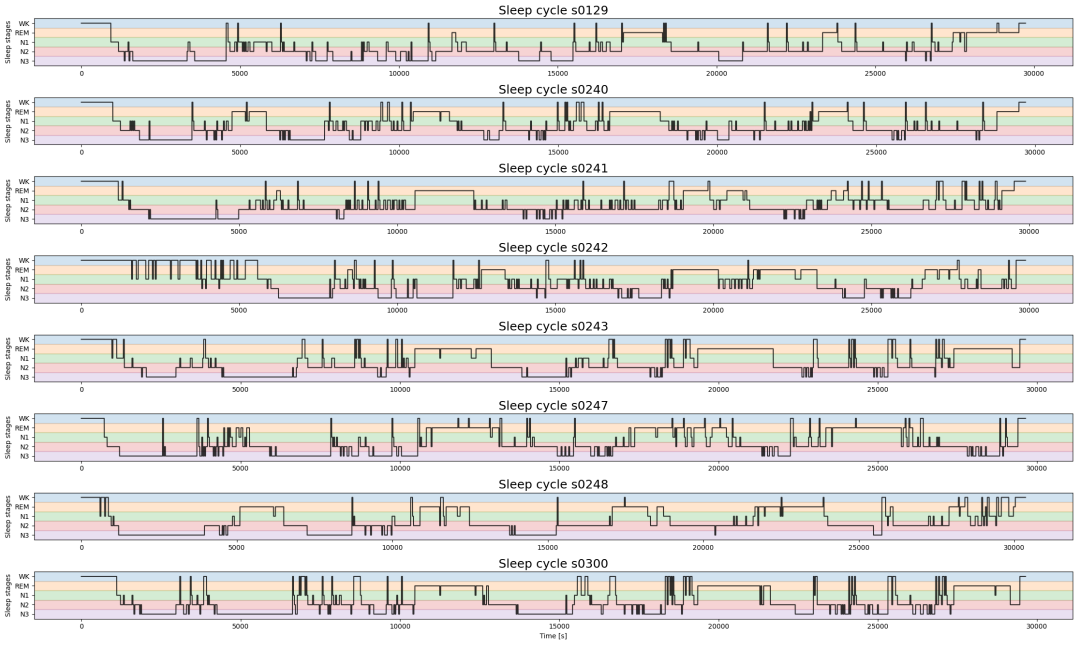


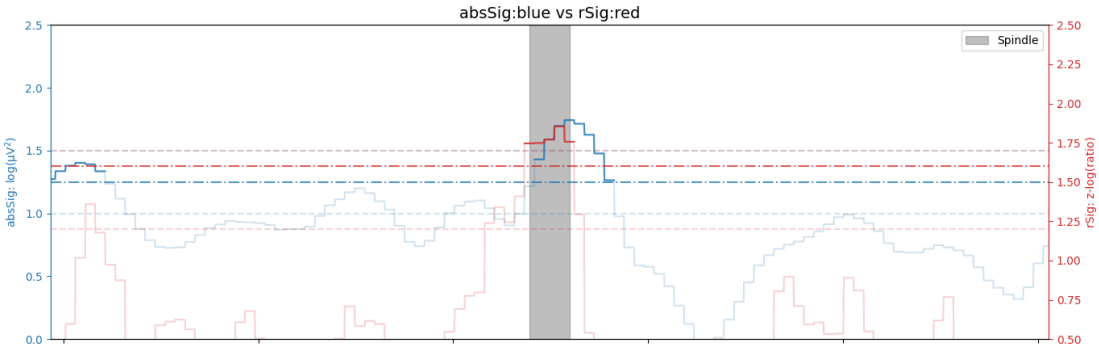
Figure 4.4: Distribution of sleep stages for all subjects

For the two subjects with expert annotation, the A7 yielded results that were less than the number of spindles from the sleep expert. The biggest variation in terms on spindles were for "Sub 0129" with a difference of 225, while the other subject had a difference of 36 spindles as seen in table (4.3).

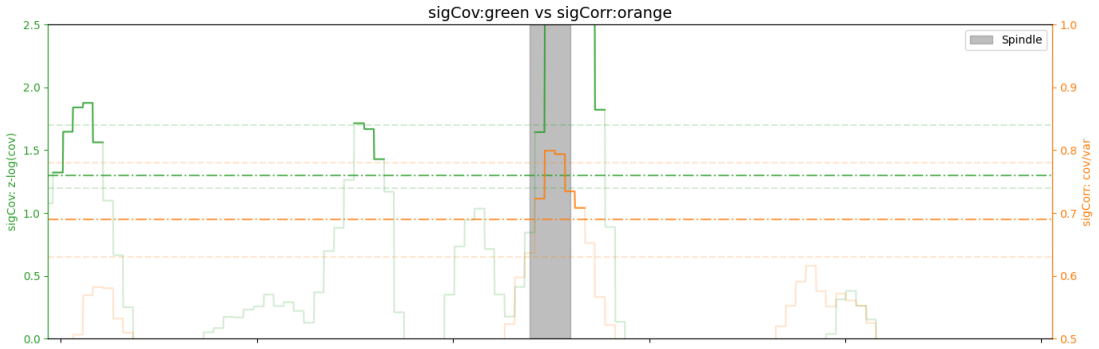
The A7 algorithm detected spindles when the four thresholds where exceeded at the same time, where these were *Absolute sigma power*, *Relative sigma power*, *Sigma covariance* and *Sigma correlation*. The calculations of these can be found in subsection 3.2.1. Figure (4.5) illustrates one spindle detection, where the spindles is indicated by the grey color. All four thresholds are exceeded at the same time within the grey area, where a darker colored graph is when the value is above the threshold. The dark dotted vertical line indicates the recommended threshold used, while the lighter ones represent the min and max thresholds that were found in the paper for the method. The spindle lasted for about 0.76 seconds.

Within this short time frame, only A7absSig and A7sigCov exceed their thresholds, while both A7rSig and A7sigCorr are below the recommended threshold used. A7rSig does exceed the minimum threshold at one point, at the same time as A7absSig and A7sigCov, but A7sigCorr is far from the minimum and therefore not detected as spindle.

To test if there were any thresholds that would produce a better overlap between the model and the expert, a visualization was done for the four thresholds at the time the expert annotated a spindle. This can be seen in figure (4.6), where the color darkens as the value of the threshold increases. The light grey, dashed vertical lines indicate the min and max thresholds, while the dark grey line indicate the recommended value. The thresholds were set according to the paper from Karine Lacourse [35], tested on several datasets to give the best results. As can be seen in the figure, there where several spindles that where located below the thresholds. This is especially apparent for the A7sigCorr-calculations, were most of spindles annotated by the sleep expert yield a value below the recommended minimum threshold. To be able to include all spindles the thresholds would need to be set at the lowest possible values. This would almost be the same as allowing all values, and annotating them to be spindles. The closeness of the min and recommended threshold for the A7sigCov, as indicated by the lines, tells that lowering this value will not include many more spindles from the expert, even though changing the other thresholds closer to minimum would imply this. Based on this information, the thresholds were set to the recommended ones from the paper.



(a) The calculations of Absolute sigma power and Relative sigma power



(b) The calculations of Sigma Covariance and Sigma Correlation

Figure 4.5: The thresholds of the A7-algorithm used to identify sleep spindles

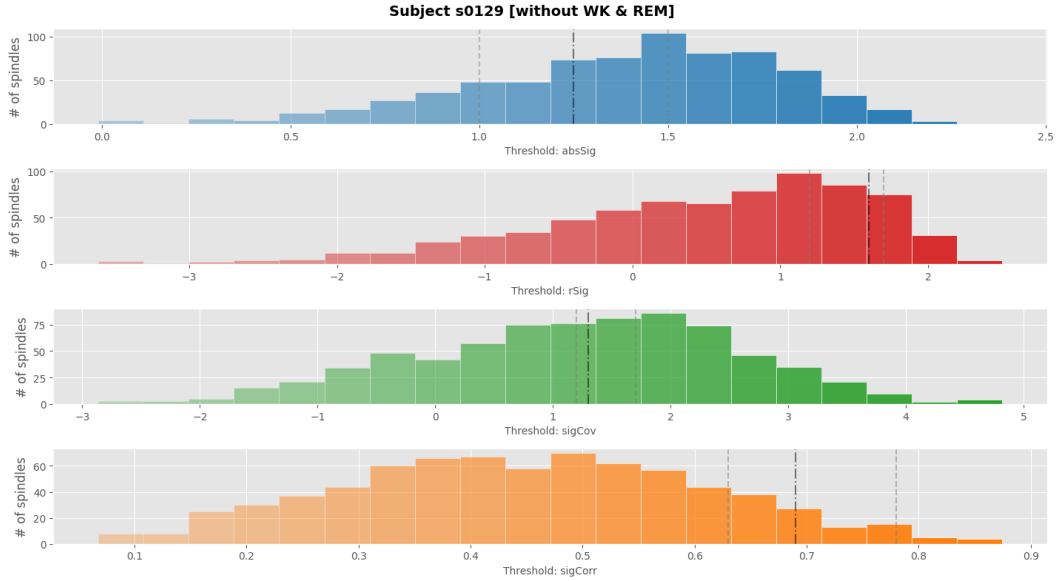


Figure 4.6: The four thresholds indicating spindles

In figure (4.7) the localization of the different spindles can be seen, with the first graph showing the annotations done by the expert, followed by the A7 algorithm before spindles that overlapped within 0.3 sec in terms of start point set by the expert. The overall interpretation is that the localization of the spindles are quite similar, with a small amount of spindles that begin at the relative same time point. The darker color indicate more spindles located in the same area, where the expert has annotated a higher occurrence relatively early in the sleep cycle, while the A7 has a bigger population of spindles located after approximately 1.5 hour (5400 sec) when returning from WK state registered in the sleep state scoring. A7 detects spindles at the first time entering

N1 sleep, while the sleep experts first spindle is at N2 sleep. Looking at the overlap, it seems to increase as time goes by, with the exception of the first detections. The last 3000 seconds of the recording there are no spindles detected by either the sleep expert or the method.

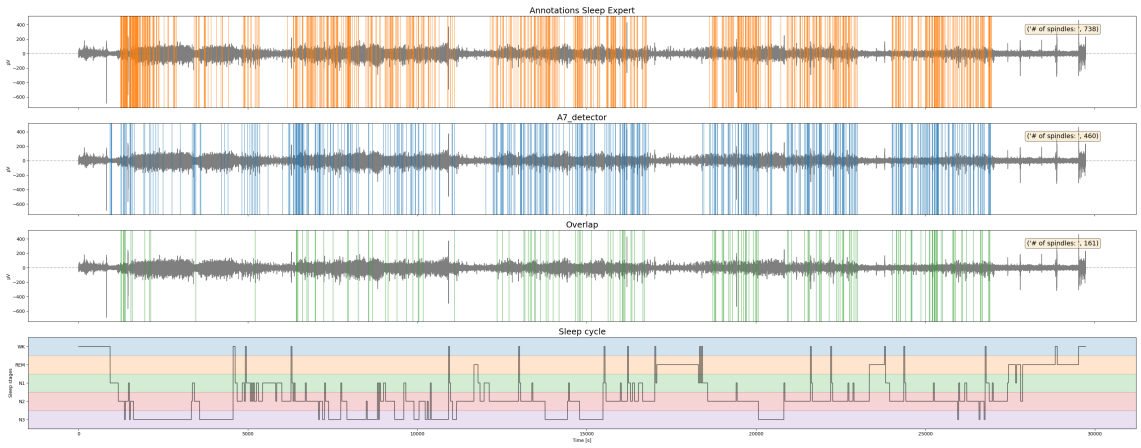


Figure 4.7: Illustration of the localization of spindles of the sleep expert, A7-algorithm and the overlap between them

The distribution of the spindles across the different sleep stages are present in fig. (4.8). The color of the spindles are reflected by the colors used in the sleep cycle plot with WK as blue and so fourth. There are two spindles reported in WK stage, while REM does not contain any spindles. The first spindle in WK is close to when the person fell asleep and entering light sleep, while the other spindle is approximately after half of the cycle has passed. The sleep cycle shows a short awakening from N2 and back to the same stage. The NREM state contain the most spindles, where N2 is prominently the one with most spindles. The spindles are recorded almost all over the sleep cycle, with a higher recurrence as time goes by. N3 sleep contains the second most spindles, with a decreasing number of detections at the end of the recording. Most spindle are present in the first half of the sleep, with one spindle standing out as one of the last spindles detected in total close to 27000 sec. The spindles in N1 are scattered at several times with a low populations for each appearance.

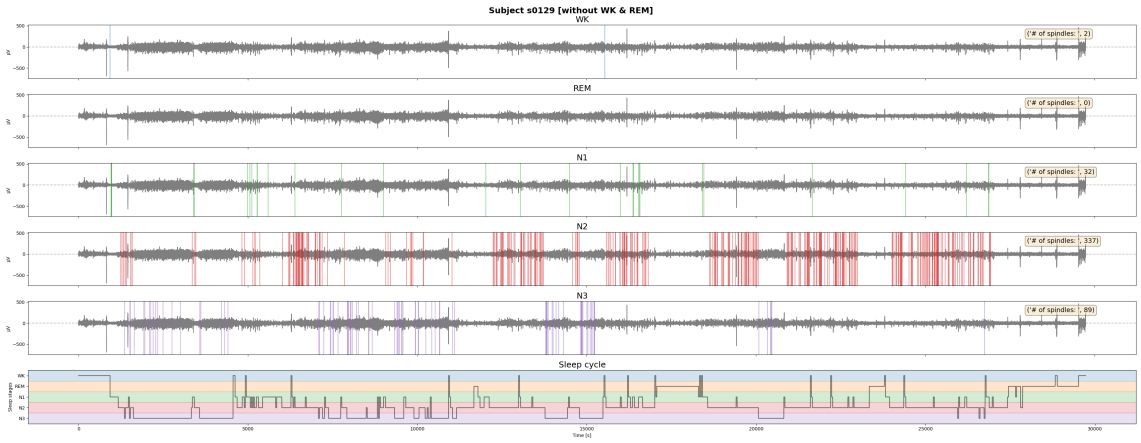


Figure 4.8: The distribution of the calculated sleep spindles, as well as the sleep cycle for the subject

4.3.3 DOSED

The DOSED algorithm requires the edf-file containing the EEG signals as well as the spindle annotations to be converted to a H5-file. This file format is one of the Hierarchical Data Formats

(HDF) used to store large amounts of data in the form of multidimensional arrays [22]. The DOSED algorithm is posted via Github, where tutorials for the converting of H5, as well as the training of dataset is supplied [59]. As already mentioned, the files needs to have the exact same number of channels, including the same names in lowercase letters to be able to run the code.

Expectations regarding number of spindles

Since two of the "Sub 0240" and "Sub 0300" only resulted in maximum of one spindle, these were used as test subjects for the DOSED method. In total the eight subjects were used, with five for training, one for validation and two subjects used for testing. The DOSED is initially tested for two channels were the dataset is resampled to 64Hz before running the code. The use of more channels is said to lead to a significant boost in detection performance, which is also stated in Chambon et al. [53]. Since this was based on a bigger dataset, four different approaches were tested to see which one would work best with the small number of subjects. The goal is not to identify absolutely every spindle in each subject, but to get the most prominent spindles to reflect a sleep expert and the basis for A7.

To get a better grip of what to expect from the DOSED and how it should be tuned to meet the goal, the study done by Purcell et al [49] were used as a foundation. The study focused on identifying characteristics of spindles as well as to calculate the expected number of spindles in sleep, based on 11.630 individuals with a age range from 4 to 97 years. They used the definitions of a spindle to be a burst of 11-15Hz activity which was typically around 0.5 sec and 2.0 sec in duration. Their result is only based on epochs of separate sleep stages, without manually annotated arousal, movements and artifacts. The dataset for N2 sleep consists of 16.499 hours of recordings, which represent 1.4 hour per individual. The average spindle density in epochs of N2 was calculated to be 1.88 per min, which is almost one spindle for every epoch of artifact-free N2 sleep. N3 sleep is usually shorter than N2, where N3 also was analysed the same way as N2 sleep, with a density of 1.45 spindle per min. The other sleep stages were not calculated the same way due to few epochs. The overall analysis showed that females had 0.16 more spindles per minute compared to males, where menstrual cycles might have impacted the spindle activity since the menstrual timing was not available information.

The calculations with the spindle density of N2 and N3 sleep were calculated for all subjects in the IIIS dataset, with the expected number of spindles can be seen in table (4.1).The calculations are based on the given average spindle density, but kept in mind that they could be 0.16 higher since the subjects were female. The calculations of number of spindles per sleep stage is found in eq. (4.1). "Sub 0240" resulted in the lowest number of expected spindles, due to the low number of epochs in N2 sleep compared to the other subjects.

The annotations from the public dataset from Montreal Archive of Sleep Studies (MASS) [8] was also used to compare expected results with the results from DOSED. This is done to be able to tune the parameters to values that would yield satisfying numbers. As seen in table (4.2), the different experts had varying annotations for the same dataset. The biggest disagreement was "Sub 1" with a deviation of 1408 spindles. The biggest agreement was "Sub 3" with 460 spindles in difference, which corresponds to the highest number of spindles the A7 algorithm found for the IIIS dataset. The MASS dataset included 19 subjects, both male and females with age 23.6 $\pm$ 3.7 [18-33] (mean $\pm$ standard deviation [range]). The PSG recordings consisted of 19 EEG channels, four EOG channels, one bipolar EMG and one ECG. The dataset are one of few datasets available online, and was used as comparison since the subject are within the same age group.

$$Number\ of\ spindle_{Subject} \approx \frac{\#\ of\ NX\ epoch}{2}\ *\ spindle\ density \tag{4.1}$$

Subject	# of N1 epochs	# of N2 epochs	# of N3 epochs	Expected # of spindles
Sub 0129	104	458	249	513
Sub 0240	139	449	158	441
Sub 0241	198	463	118	422
Sub 0242	193	262	204	338
Sub 0243	85	409	187	433
Sub 0247	55	351	277	455
Sub 0248	74	437	211	470
Sub 0300	58	323	295	448

Table 4.1: Expected number of spindles for IIIS dataset, based on from 11.630 individuals with age 4 to 97

Subject	E1	E2	Difference E1 & E2
Sub 1	1044	2452	1408
Sub 2	1143	2212	1069
Sub 3	143	603	460
Sub 4	253	-	-
Sub 5	341	1201	860
Sub 6	150	841	691
Sub 7	913	1608	695
Sub 8	385	-	-
Sub 9	814	1670	856
Sub 10	795	1939	1144
Sub 11	606	-	-
Sub 12	709	1204	495
Sub 13	698	1438	740
Sub 14	713	1618	905
Sub 15	97	-	-
Sub 16	453	-	-
Sub 17	470	1192	722
Sub 18	1164	1680	516
Sub 19	315	1058	743

Table 4.2: Number of spindles detected by two different sleep experts as well as the difference between them (Source: [[8]])

Predictions done by DOSED

The distribution of spindles in all stages for "Sub 0129" in fig. (4.7) based on A7, together with the calculations based on epochs as well as the two experts annotations from MASS dataset gives an interpretation and expectation of what one could expect when running DOSED. The parameters were tuned to meet the expected result, and therefore set very strict to limit the number of default events not matching any true events since is the disadvantage of the model.

The results of the four approaches of DOSED are presented in tab. (4.3), with differentiating the sampling frequency as well as number of channels used. "Sub 0240" and "Sub 0300" where the test subjects, since they had performed the worst with the A7-method and in theory did not contain any annotations. The same parameters were tested for all approaches, and had been tuned to appropriate the expected number of spindles.

For "Sub 0240" the spindle annotation range from 2384 spindles and till 47514 spindles, depending on the sampling frequency and number of EEG channels. The lowest number of spindles is for all channels with a sampling frequency of 256Hz, while the highest is for the same sampling frequency with only two channels.

As the table shows for "Sub 0300", the result ranged from 262 spindles to 26211 spindles, where the last is almost 100 times as many spindles, and at least ten times higher than what one could expect

Subject	Expert	A7	DOSED			
			64Hz		*256Hz*	
			2 EEG channels	*128 EEG channels*	*2 EEG channels*	*128 EEG channels*
Sub 0129	738	460	-	-	-	-
Sub 0240	X	0*	2967	7550	47514	2384
Sub 0241	X	254	-	-	-	-
Sub 0242	204	168	-	-	-	-
Sub 0243	X	150	-	-	-	-
Sub 0247	X	299	-	-	-	-
Sub 0248	X	6	-	-	-	-
Sub 0300	X	1**	2960	6384	26211	261

Table 4.3: Number of spindles detected by sleep expert as well as the two models A7 and DOSED.
* *Same result after six rounds of cleaning*
** *Needed more than two rounds of cleaning*

according to the experts from MASS dataset and their range of annotations. Using a sampling frequency of 256Hz resulted in both the highest and lowest number, where in this case applying all channels would give the most comparable result based on the above mentioned methods. Fig. (4.9) shows all the detected spindles and their location. The spindles are mostly localized around 3-4 hours after the participant went to sleep. Comparing with the visualization of the epochs annotation, the spindles detected are in REM, N2 and N3 sleep.

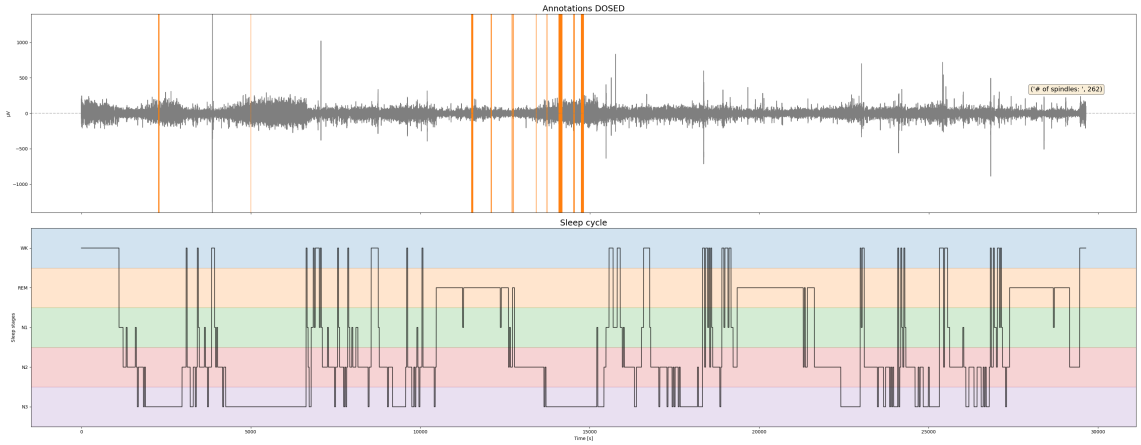


Figure 4.9: All spindles detected with 256Hz and 128 EEG channels for "Sub 0300"

Looking at figure (4.10), the distribution of spindles are found in three stages, where N3 stands out as the stage with the most spindles. REM is the second stage with 53 spindles, and the stage with the least detected spindles out of the three are stage N2.

4.4 Cost effective and impartial

The goal of using the spindle detectors in a different way then how the papers present them, where bigger datasets are created and compared with manual annotated information by experts. This idea is based on using the detectors in a new way, to utilize the resources in for instance a lab or research center in a better way. The result would also be less dependent on the individual interpretations, which are clear that influence the way spindles are detected, and the characteristics that are fundamental.

The time spent annotating sleep scoring and spindles for one subject is very dependent on the person doing the scoring as well as the dataset. A lot of experience may lead a scorer to complete the spindle annotations up to about two hours, while someone new might use several hours on the

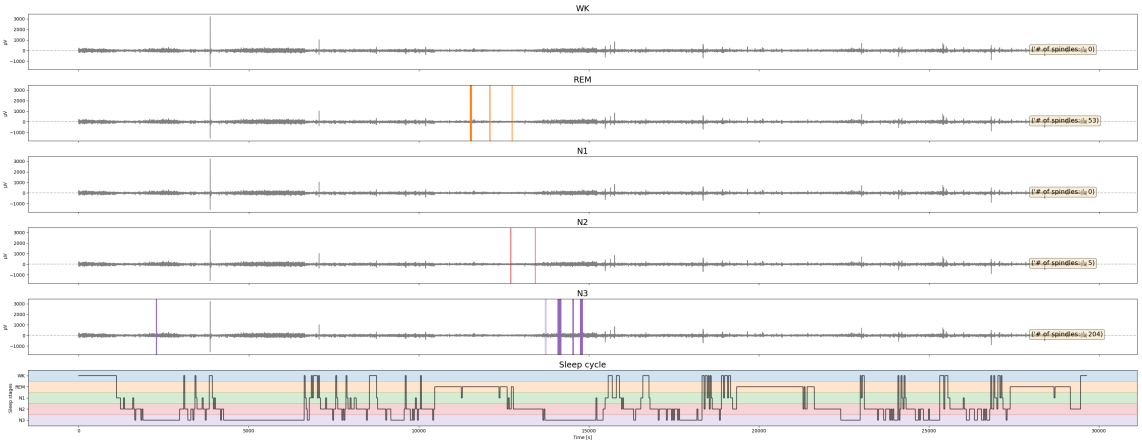


Figure 4.10: Distribution of spindles over the different stages detected with 256Hz and 128 EEG channels for "Sub 0300"

same dataset. The time spent for annotations of data has not been thoroughly disclosed, and this report have therefore set the time spent for scoring and spindle annotation to be approximately five hours, where the sleep scorer did both annotations at the same time and therefore not referred to previous sleep scores to eliminate selected epochs before or after the person clearly is awake.

Figure (4.11) illustrates the time used to build up a database, where all data will need to go through several stages before it can be used in the prediction algorithm. It will therefore not save much time in the total process, but much of the work for the sleep expert may now be moved to another person capable of doing the preprocessing as well as run the A7 and DOSED.

In the figure the first bar indicates the time spent for a sleep scorer to complete the manual scoring and detection, where the sleep expert will not be able to work on any other matters simultaneously. There is nothing automatized, and each epoch is visually inspected for spindles as well as defined according to the characteristics of the sleep stages. This continues until the recordings is finished or a person is confirmed to have waken up and the rest of the recordings will not be useful for sleep research. Since sleep research often requires an adaption night, where this recording might be useful to analyze to make sure the subject does not have any abnormalities in their cycle that might inflect the purpose of the recording. This will result in at least two recording per subject for those that are accepted into the study.

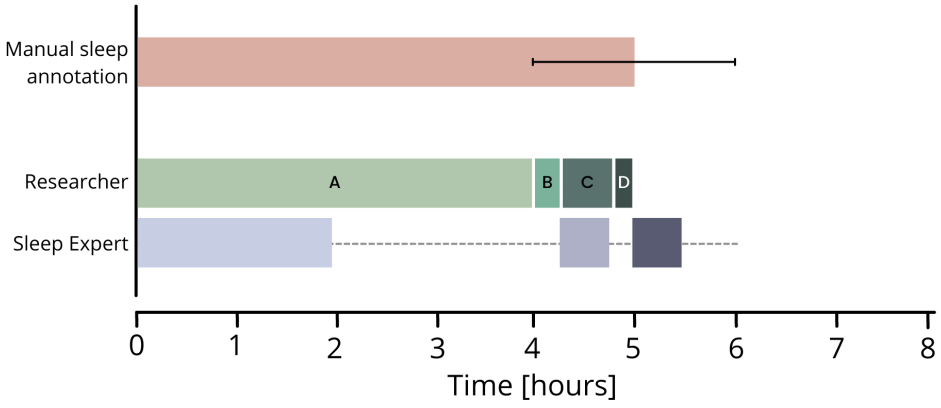


Figure 4.11: Time spent for sleep scoring and spindle annotation for one subject

The next two bars indicate approximate times used by researcher and the sleep expert for the IIIS dataset. The bar labeled "researcher" contains all work with the data after sleep staging was completed, and includes all requirements to run A7 as well as DOSED. The time the sleep expert used for the same subject is indicated by the bar below.

The first area of the bar, the light green color and labeled "A", indicates the time spent for preprocessing. This is without an doubt the most time-consuming part where the fitting of ICA took the longest time. To create 40-50 components based on 128 EEG channels, one had to wait for approximately 40 min. After removing the components, the sensor signals will need to be reconstructed. To be able to make sure the most prominent artifacts are removed, it is recommended to create new components, and possible remove more artifacts before reconstructing and finish the signal processing. Running two rounds of ICA took approximately four hours from the raw signal was opened and until it was reconstructed after artifact removal.

The running of A7-algorithm took approximately ten minutes and is indicated by label "B" in the bar graph for the researcher, with five minutes added to create the required files needed to calculate the thresholds and detect spindles.

Both "C" and "D" are steps of the DOSED method, where the first is to create the H5-files with the data and annotations. Depending on the size of the annotations and number of channels, the average time for this was roughly 30 min. "D" represents the run time for the DOSED which was approximately 10 min for one subject. In total the time for this approach would take as long time as the manual annotation.

For the sleep expert in fig. (4.11), the first bar indicates the two hours spent to score the sleep stages for the subject. The next two bars are placed according to when the sleep expert is needed to verify results and give feedback regarding tuning. After the A7 calculations the sleep expert verified that the results looked satisfying, and approved the dataset for further analysis with machine learning. The last bar indicates the verification needed to look at for instance 10% of the data with annotated spindles. This should be sufficient to give feedback on the results before potential re-tuning.

However will the time spent for spindle detection change drastically when there is enough data to work as the basis, and then add unknown data to it for further predictions. This can be seen in fig. (4.12), where "Researcher" has been replaced by "Automatic detection". The difference is now that there will is only needed to run the DOSED method, to predict spindles. In the bar labeled "Automatic detection", the first area called "I" represents an automatic sleep scoring method. There are many found online to test, which will annotate accurately based on the AASM manual and characteristics given. Since this was not utilized in this report there is not an accurate time for this, but for one subject this should run quite fast and is here approximated to last for 30min. From this the unknown dataset will need to be converted to an H5-file to work with DOSED, before it is predicted by the DOSED. The sleep expert will only focus on verification of results, and will therefor only focus on for instance 10% of the data.

"II" represents the converting of H5-files, where this only would apply to new datasets used in DOSED. The "III" bar is the running of DOSED, where this will require more time as the database increases. However it will not need to monitored to be able to run.

For the sleep expert their knowledge about sleep will only be needed to look at a selected extract of the annotations and data. Comparing to the manual annotations previously required, their time may be limited to only 30 min for verification which is a great improvement and use of resources when other options are available.

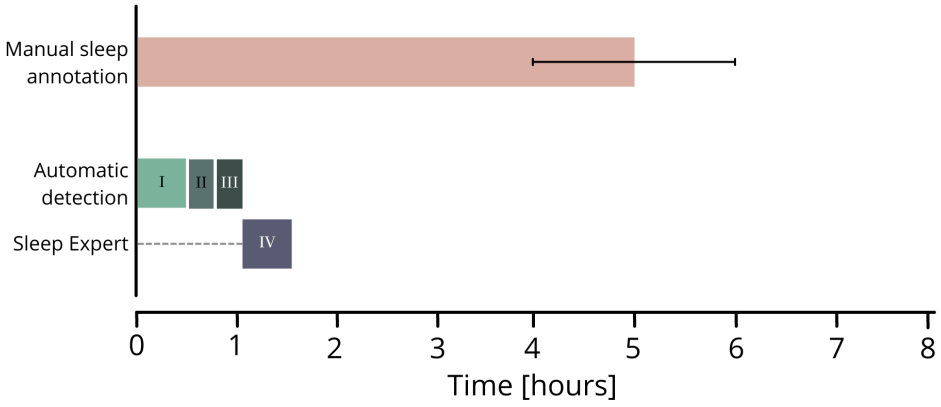


Figure 4.12: Time spent for sleep scoring and spindle annotation after the basis is created

Chapter 5

Conclusion

This project focused on how to be able to build up a detection database with minimal need for sleep expert to limit their workload. Data containing a lot of artifacts will not be well predicted when a database containing artifact-free data. To be able to save time in the future, a new database will need to be created where this would imply many hours for the sleep expert. However are the automatic detection methods available online, and constantly changing to improve the understanding of spindles, as well as emulate the sleep expert.

5.1 Summary and conclusion

By utilizing the different software detectors one will be able to build up a database specialized for a specific lab, and therefore be able to detect artifacts that might be caused by equipment complications or geographical based due to power lines or other electrical objects. Preprocessing data is a time-consuming work, with many ways to approach the problem.

5.1.1 Dataset

Data quality and availability is crucial for machine learning, where it is impossible to configure algorithms to control the flow of low-quality and inaccurate data. This is a significant challenge for this project, where the dataset used is very small. However does it show promise, and shows that there are possibilities to utilize detectors in a new way, where they can replace tedious workload to gain more consistent results.

Since some of the data were collected from a study of menstrual cycle, this could have affected the results. The study by S. M. Purcell et al. [49] stated that menstrual cycles could impact spindle activity, but due to no information on the matter, they could neither confirm nor deny the matter. The spindle density between females and males were noticeable, and needs to be taken into consideration. One of the other studies were based on simultaneously measurements from PSG as well as a portable EEG device. Whether or not this included more artifacts are difficult to conclude since there is no given information about which study the subjects originate from.

When removing artifacts, "Sub0248" showed channel noise at channel A1. This was removed, where new rounds of ICA did not display the similar component, and should not have interfered with the results. This channel was only used for predictions that utilized all channels, and not for single channel analysis. "Sub0247" also showed channel noise from channel A2, where the same way of thinking was applied as previously mentioned.

"Sub 0240" did have a sleep cycle that was unusual with a uniformly distribution for all sleep stages. The subject also needed a lot more time before entering deep sleep, compared to the other cycles. The subject showed satisfying spindle detection with A7, and since the number of subjects

were low the data was not excluded.

5.1.2 A7

The performance of A7 were quite accurate to what the expert annotated, as well as gave results that would correspond to what one could expect. However since it required artifact-free data, the preprocessing in advance demanded much time, in contrast to the fast run time for the algorithm. The overlap between the expert and the algorithm were approximately 22% for "Sub 0129" based on when the spindles were set to begin with a range for $\pm$ 0.3 sec to not be too strict on exact match. The distribution of spindles as well as the population of them where however satisfactory to the sleep expert.

Inaccurate cleaning of data was however a disadvantage, and for three of the subjects a satisfying calculation of spindles was not obtained. Even though the process was the same for all subjects, as well as the results for each round of cleaning were tested in the A7. Large spikes are present in the representation of spindles for "Sub 0300", and the signal-to-noise ratio was quite low for all subjects.

To be able to utilize this method to its fullest, there needs to be more focus on the cleaning of data, and removing of artifacts. The A7 is initially developed to detect spindles for N2, but did give satisfying results for the other stages as well when removing WK and REM. However did the A7 detect two spindles in WK for "Sub 0129", even though these epochs were marked as artifact and therefore not suppose to be involved in the identification. The spindles were close to the subject entering N1 sleep, and therefore not detected in a period with a long WK period. The reasons behind these detections are unclear, but due to the low number of detected spindles for the two sleep stages no further actions were acquired.

5.1.3 DOSED

The DOSED required data with annotations to be able to predict for unknown subjects, where parameter tuning and getting a good enough basis are the great challenges for machine prediction methods. Due to the data having artifacts, many of the available artifact-free dataset did not predict good enough on the unknown participants chosen from the IIIS dataset. It was clear that eight subjects were not sufficient, and more data gathered from the same lab could yield better results. However does it show potential and will with time and more training give better predictions.

There are many detectors based on machine learning, where the biggest challenge is the gathering of data. There needs to be a good distribution, as well as a diverse dataset. The subjects were all females withing the same age range, and therefore a has low variation. Using more channels indicated a better result with the low number of subjects, were a bigger dataset should work well with only a few channels. The great advantage of using DOSED is its ability to add more events to predict, even though this was not practised for this project.

The DOSED method did not require score staging, which also may have impacted the results. As previously mentioned, there are bigger differences among scorers how to label N1 and N3 sleep, compared to REM, N2 and WK. The DOSED algorithm looks independently on the data, not taking into account which stage is present. For the A7 detection only NREM were focused on, since REM and WK could have characteristics that could simulate sleep spindles. Compared to the total number of spindles detected, approximately 20% were found at sleep labeled as REM by the expert.

5.1.4 Semiautomatic spindle detection platform

With minimal sleep expert intervention, the semiautomatic spindle detection platform may be a new way of thinking which should be explored more. Most models found online either requires a

large database and/or annotated data to be able to detect spindles. Or they are only able to label spindles for one recording at a time, where tedious preprocessing is required.

However shows the platform promise with only a small dataset, where the ambition would be a great support for sleep experts. Mainly it would be used as the only detection method for spindle, although it would also work as support for manual identifications used for teaching purposes. Using more automatic detectors may also result in better definitions of characteristics for events, which will again lead to more accurate identification for spindles. The compatibility should increase with automatic detectors due to the removing of subjective annotation caused by human error and interpretation.

The greatest motivation is still to remove time-consuming and repetitive work for sleep experts, so they could better use their time instead of being stuck with tiresome tasks.

5.2 Discussion

Working with a dataset containing a lot of artifacts require a lot of preprocessing, where the operation is currently not optimized and require visual inspection and selection. Out of the entire project, this has been the biggest challenge. The work was repetitive and monotonous, and will be prone to subjective bias. Funny enough this is the exact same descriptions that was the motivation to change the way to do spindle annotation. Artifacts appear as a result of both things a person may control, such as equipment and relaxing environment, as well as uncontrollable things, like geographical placement and electronic devices. Being able to minimize these impacts should result into a better database, and is a motivation to take precaution. Further analysis of the data will never flourish if the basis is corrupted, and the results will represent the way it was collected. Being able to automate and optimize this process would go a long way to minimize workload.

The size of the data also makes it exposed and easily disturbed by small changes, since the model is still trying to learn how the event may appear. It is not able to get a good interpretation of what to look for, and will most likely annotate a high number of false positives.

Due to the big difference when referencing to the scoring of the MASS data, a long side other studies based on a bigger group annotating the same data, show a big inconsistency in the labeled data. This may be the cause of the standard manuals not being clear enough, which again will cause in accurate basis when used in automatic detections. However should a big database be able to handle some deviations, but is also a big motivation in automatising more of the steps in sleep research. Rosenberg R.S. et al. [51] addresses that the AASM manual was not satisfactory for transitions between sleep stages, when comparing several sleep experts annotations, in addition to scoring of stages N1 and N3.

However is the semiautomatic spindle detection something to continue explore to minimize the intervention required by a sleep expert. This could be to look into other detectors that could give more accurate results, as new models and detectors are introduced and improved. The development of algorithms requiring a small amount of data would be a solution and great tool in emulating sleep experts. However, until this becomes a possibility the focus should initially be to gather data to incorporate in the database.

5.3 Future work

Further work with new subjects will be required if one wants to be able to build up a database able to handle artifact data, so that there is minimal interference and more time sufficient.

For many of the papers read when working on this project, as well as the different detectors available for testing, are using different definitions for spindles either regarding duration or frequency compared to AASM. Most are using the most common definition for spindles, which represents the fast spindle, as well as accepting shorter duration. This will all inflict the automatic detect-

ors, when there is no standard way used in practice. Further work with getting a more accurate spindle definition so that the reproducibility increases and that there is no uncertainty in what specifications the given paper or research team has used for their application.

With the development happening with machine learning and automatic detection in several fields, other models are definitely something to look into if they can solve any of the challenges involving the bias. Most of the database created are based on manual annotation, and a machine learning algorithm will therefore be prone to bias from the beginning. This should be the biggest motivation to create an official test database, where other labs can incorporate their data to further build upon. The data needs to fulfill the requirements for a good database, with diversity and accurate definitions.

Other software detectors could either work as a supplement or replacement for the already mentioned and tested software. Getting more accurate and precise models will only be relevant if they can handle a small database. Nevertheless, the advantage of DOSED is its possibility to look for several events, and therefore could be highly useful for detection micro-architecture events.

The pipeline before entering DOSED may not be from A7, but the way of thinking should apply for other events. However should DOSED be altered to able to handle data with different amount of channels. A more dynamic selection of what channels to use in the predictions would also be a great advancement and increase its usability, with a limitation that the channels need to be the same for all data.

Abbreviations

A7 Lacourse Spindle Detector

AASM American Academy of Sleep Medicine

AP Action Potential

CNS Central Nervous System

CPU Central Processing Unit

DOSED Dreem One Shot Event Detector

ECG Electrocardiogram

ECoG Electrocorticography

EEG Electroencephalogram

EEGbf Band-passed filtered EEG (0.3-30Hz)

EMG Electromyographic

EOG Electrooculogram

ERP Event-Related Potential

GPUs Graphics Processing Units

HDF Hierarchical Data Formats

ICA Independent Component Analysis

IIIS International Institute for Integrative Sleep Medicine

IoU Intersection over Union

MASS Montreal Archive of Sleep Studies

MEG Magnetoencephalography

NREM Non-rapid Eye Movement

PC Principal Component

PSA Power Spectral Analysis

PSG Polysomnography

REM Rapid Eye Movement

SD Standard Devitation

SO Slow Oscillation

Bibliography

[1] AASM. *AASM releases updated version of sleep scoring manual*. URL: https://aasm.org/aasm-updated-version-sleep-scoring-manual/ (visited on 17/11/2022).

[2] AASM. *American Academy of Sleep Medicine*. URL: https://aasm.org/clinical-resources/scoring-manual/ (visited on 17/11/2022).

[3] E. Newton Harvey Alfred L. Loomis and Garret A. Hobart. 'Distribution of distrubance-patterns in the human electroencephalogram, with special reference to sleep'. In: *Journal of Neurophysiology* 1 (1938), pp. 413–430.

[4] Minnesota Brain Injury Alliance. *About Brain Injury: A Guide to Brain Anatomy*. URL: https://www.braininjurymn.org/library/AGuideToBrainAnatomy.pdf (visited on 29/11/2022).

[5] tutor2u AQA A Level Psychology Course Companion (Edition 1). *Studying The Brain*. URL: https://www.thomasclarksonacademy.org/attachments/download.asp?file=516&type=pdf (visited on 05/12/2021).

[6] MD Brendon Peters. *Overview of Hypnagogic Hallucinations*. URL: https://www.verywellhealth.com/hypnagogic-hallucination-3014888 (visited on 28/11/2022).

[7] Shure C. *The Forgotten History of Alpha Brain Waves*. URL: https://www.bitbrain.com/blog/alpha-brain-waves (visited on 13/12/2022).

[8] CEAMS. *SS2 Biosignals and Sleep stage*. https://borealisdata.ca/dataset.xhtml?persistentId=doi:10.5683/SP3/K26LXJ. 2022. DOI: 10.5683/SP3/K26LXJ.

[9] Manish Chablani. *YOLO - You Only Look Once, real time object detection explained*. URL: https://towardsdatascience.com/yolo-you-only-look-once-real-time-object-detection-explained-492dc9230006 (visited on 11/12/2022).

[10] Arnal P J Chambon S Thorey V. *DOSED: A deep learning approach to detect multiple sleep micro-events in EEG signal*. https://pubmed.ncbi.nlm.nih.gov/30946878/. 2019. DOI: 10.1016/j.jneumeth.2019.03.017.

[11] Kenda Cherry. *The 4 Stages of Sleep*. URL: https://www.verywellhealth.com/the-four-stages-of-sleep-2795920 (visited on 28/11/2022).

[12] Timplalexis Christos. *Classification of Sleep Stages Using Machine Learning*. URL: https://repository.ihu.edu.gr/xmlui/bitstream/handle/11544/29403/Classification%5C%20of%5C%20Sleep%5C%20Stages%5C%20Using%5C%20Machine%5C%20Learning%5C%20Methods.pdf?sequence=1 (visited on 08/12/2022).

[13] Kanwaljit Singh Danielle Brigham Yash Shah. *Comparison of artifacts between paste and collodion method of electrode application in pediatric EEG*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6931097/. 2020. DOI: 10.1016/j.cnp.2019.11.002.

[14] Zeitlhofer J Danker-Hopfe H Anderer P. *Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard*. https://pubmed.ncbi.nlm.nih.gov/19250176/. 2009. DOI: 10.1111/j.1365-2869.2008.00700.x.

[15] UC San Diego David Bailot. *Illustrations of neuron*. URL: https://www.analytica-world.com/en/news/1156404/why-are-neuron-axons-long-and-spindly.html (visited on 13/12/2022).

[16] Ben Dickson. *Why data remains the greatest challenge for machine learning projects*. URL: https://venturebeat.com/ai/why-data-remains-the-greatest-challenge-for-machine-learning-projects/ (visited on 10/12/2022).

[17] EEGLAB. *Independent Component Analysis for artifact removal*. URL: https://eeglab.org/tutorials/06_RejectArtifacts/RunICA.html (visited on 09/12/2022).

[18] Dr. Nilong Vyas Eric Suni. *Stages of sleep*. URL: https://www.sleepfoundation.org/stages-of-sleep (visited on 28/11/2022).

[19] Haas L. F. 'Hans Berger (1873-1941), Richard Caton (1842-1926), and electroencephalography'. In: *Journal of neurology, neurosuergery and psychiatry* 74 (2003), pp. 9–9.

[20] Murray M.M Franceschiello B. Biasiucci A. 'Electroencephalography'. In: *Current Biology* 29 (2019), pp. 80–85.

[21] Larson E Gramfort A Luessi M. 'MEG and EEG Data Analysis with MNE - Python'. In: *Frontiers in Neuroscience* 7 (2013), pp. 1–13.

[22] The HDF Group. *Hierarchical Data Format 5 File*. URL: https://fileinfo.com/extension/h5 (visited on 17/12/2022).

[23] Schulz H. *Rethinking sleep analysis*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2335403/. 2008.

[24] Kate Hoy. *Explainer: the brain*. URL: https://theconversation.com/explainer-the-brain-11196 (visited on 13/12/2022).

[25] Queensland Brain Institute. *Corpus callosum*. URL: https://qbi.uq.edu.au/brain/brain-anatomy/corpus-callosum (visited on 29/11/2022).

[26] Queensland Brain Institute. *Lobes of the brain*. URL: https://qbi.uq.edu.au/brain/brain-anatomy/lobes-brain (visited on 29/11/2022).

[27] BioSemi Instruments. *Headcaps*. URL: https://www.biosemi.com/headcap.htm (visited on 17/11/2022).

[28] *International Institute for Integrative Sleep Medicine*. URL: https://wpi-iiis.tsukuba.ac.jp/ (visited on 11/11/2022).

[29] Begoña Garcia-Zapirain Jose Antonio Urigüen. *EEG artifact removal—state-of-the-art and guidelines*. https://iopscience.iop.org/article/10.1088/1741-2560/12/3/031001/meta. 2015. DOI: 10.1088/1741-2560/12/3/031001.

[30] Dries De Kegel. *Illustrations of the layers of brain and head*. URL: https://www.researchgate.net/figure/Anatomy-of-the-human-head-showing-the-scalp-skull-meninges-and-the-brain-adapted_fig1_334898971 (visited on 13/12/2022).

[31] Nihon Kohden. *Polysmith Sleep Systems*. URL: https://eu.nihonkohden.com/en/products/neurology/polysmithsleepsystems.html (visited on 07/12/2022).

[32] Nihon Kohden. *PSG-1100*. URL: https://us.nihonkohden.com/products/psg-1100/ (visited on 07/12/2022).

[33] Anthony L. Komaroff. *Are toxins flushed out of the brain during sleep?* URL: https://www.health.harvard.edu/mind-and-mood/are-toxins-flushed-out-of-the-brain-during-sleep (visited on 16/12/2022).

[34] ICLabel Tutorial: EEG Independent Component Labeling. *Telling Components Apart*. URL: https://labeling.ucsd.edu/tutorial/labels (visited on 16/12/2022).

[35] Beaudry J Lacourse K Delfrate J. *A sleep spindle detection algorithm that emulates human expert spindle scoring*. https://pubmed.ncbi.nlm.nih.gov/30107208/. 2010. DOI: 10.1016/j.jneumeth.2018.08.014.

[36] Dalal SS Leske S. *Reducing power line noise in EEG and MEG data via spectrum interpolation*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6456018/. 2019. DOI: 10.1016/j.neuroimage.2019.01.026.

[37] P. J. Beadle Lisha Sun Ying Liu. *Independent component analysis of EEG signals*. https://ieeexplore.ieee.org/document/1504590. 2005. DOI: 10.1109/IWVDVT.2005.1504590.

[38] Gunnar Tofte Magnus Själander Magnus Jahre. *IDUN*. URL: https://www.hpc.ntnu.no/idun/ (visited on 10/12/2022).

[39] Ashley Marcin. *What Is Deep Sleep and Why Is It Important?* URL: https://www.healthline.com/health/deep-sleep#sleepstages (visited on 16/12/2022).

[40] Johns Hopkins Medicine. *Electroencephalogram (EEG)*. URL: https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/electroencephalogram-eeg (visited on 05/12/2022).

[41] MNE-Python. *Filtering and resampling data*. URL: https://mne.tools/0.15/auto_tutorials/plot_artifacts_correction_filtering.html (visited on 09/12/2022).

[42] MNE-Python. *find$_b$ads$_e$og*. URL: https://mne.tools/0.15/generated/mne.preprocessing.ICA.html#mne.preprocessing.ICA.find_bads_eog (visited on 09/12/2022).

[43] MNE-Python. *mne.filter.notch$_f$ilter*. URL: https://mne.tools/stable/generated/mne.filter.notch_filter.html (visited on 09/12/2022).

[44] MNE-Python. *mne.filter.resample*. URL: https://mne.tools/stable/generated/mne.filter.resample.html (visited on 09/12/2022).

[45] MNE-Python. *Repairing artifacts with ICA*. URL: https://mne.tools/stable/auto_tutorials/preprocessing/40_artifact_correction_ica.html (visited on 09/12/2022).

[46] Marshall L Mölle M Bergmann TO. 'Fast and slow spindles during the sleep slow oscillation: disparate coalescence and engagement in memory processing'. In: *Sleep* 10 (2011), pp. 1411–1421.

[47] Jonathan Godbout Nicolas Martin Marjolaine Lafortune. 'Topography of age-related changes in sleep spindles'. In: *Neurobiology of Aging* 34 (2013), pp. 468–476.

[48] Nielsen T O'Reilly C. *Automatic sleep spindle detection: benchmarking with fine temporal resolution using open science tools*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4478395/. 2015. DOI: 10.3389/fnhum.2015.00353.

[49] Demanuele C. Purcell S. Manoach D. *Characterizing sleep spindles in 11,630 individuals from the National Sleep Research Resource*. https://www.nature.com/articles/ncomms15930. 2017. DOI: 10.1038/ncomms15930.

[50] Gorski P Rejer I. *Benefits of ICA in the Case of a Few Channel EEG*. https://pubmed.ncbi.nlm.nih.gov/26738010/. 2015. DOI: 10.1109/EMBC.2015.7320110.

[51] Van Hout S Rosenberg RS. *The American Academy of Sleep Medicine inter-scorer reliability program: sleep stage scoring*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3525994/. 2013. DOI: 10.5664/jcsm.2350.

[52] Downey R Rundo JV. *Polysomnography*. https://pubmed.ncbi.nlm.nih.gov/31277862/. 2019. DOI: 10.1016/B978-0-444-64032-1.00025-4.

[53] P. J. Arnal S. Chambon M. N. Galtier. 'A Deep Learning Architecture for Temporal Sleep Stage Classification Using Multivariate and Multimodal Time Series'. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 26 (2018), pp. 756–769.

[54] Ram Sagar. *What is SSD & How It Improved Computer Vision Forever*. URL: https://analyticsindiamag.com/what-is-ssd-how-it-improved-computer-vision-forever/ (visited on 11/12/2022).

[55] J Sanei & Chambers. *EEG signal processing*. John Wiley & Sons, 2007.

[56] Stephen H. Sheldon. *Development of Sleep in Infants*. Principles and Practice of Pediatric Sleep Medicine (Second Edition), 2014.

[57] Saadat M Shrivastava D Jung S. *How to interpret the results of a sleep study*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4246141/. 2015. DOI: 10.3402/jchimp.v4.24983.

[58] Mayo Clinical Staff. *EEG (electroenchephalogram)*. URL: https://www.mayoclinic.org/tests-procedures/eeg/about/pac-20393875 (visited on 16/12/2022).

[59] Pierrick J. Stanislas Chambon Valentin Thorey. *Dreem One Shot Event Detector*. URL: https://github.com/Dreem-Organization/dosed (visited on 15/11/2022).

[60] Jan Born Til O. Bergmann. 'Phase-Amplitude Coupling: A General Mechanism for Memory Processing and Synaptic Plasticity?' In: *Neuron* 97 (2018), pp. 10–13.

[61] Chow TW. 'Personality in frontal lobe disorders'. In: *Curr Psychiatry Rep* 5 (2000), pp. 446–451.

[62] Hurley DA Van de Water AT Holmes A. *Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography–a systematic review*. https://pubmed.ncbi.nlm.nih.gov/20374444/. 2019. DOI: 10.1111/j.1365-2869.2009.00814.x.

[63] Welinder P Warby SC Wendt SL. 'Sleep-spindle detection: crowdsourcing and evaluating performance of experts, non-experts and automated methods'. In: *Nat Methods* 4 (2014), pp. 385–392.

[64] Schröger E Widmann A. *Filter effects and filter artifacts in the analysis of electrophysiological data*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3391960/. 2012. DOI: 10.3389/fpsyg.2012.00233.

[65] G.F. Woodman. 'A brief introduction to the use of event-related potentials in studies of perception and attention'. In: *Attention, Perception, and Psychophysics* 72 (2010), pp. 2021–2046.

[66] Makoto Satoh Yoko Suzuki Fusae Kawana and Takashi Abe. 'The abrupt shift to slower frequencies after arousal from sleep in healthy young adults'. In: *Journal of Clinical Sleep Medicine* 48 (2021), pp. 387–411.

[67] Li H Zhang H Wang X. *Auto-annotating sleep stages based on polysomnographic data*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8767308/. 2021. DOI: 10.1016/j.patter.2021.100371.

List of Figures

List of Tables

Appendix

A SEMIAUTOMATIC SPINDLE DETECTION PLATFORM WITH MINIMAL EXPERT INTERVENTION BASED ON INTERACTION BETWEEN TWO SOFTWARE

Susann Masdal[1], *Yoko Suzuki*[2], *Junya Furuki*[2], *Takashi Abe*[2], *Marta Molinas*[1]
Department of Engineering Cybernetics, Norwegian University of Science and Technology, Trondheim, Norway[1]
International Institute for Integrative Sleep Medicine (WPI-IIIS), University of Tsukuba, Tsukuba, Japan [2]

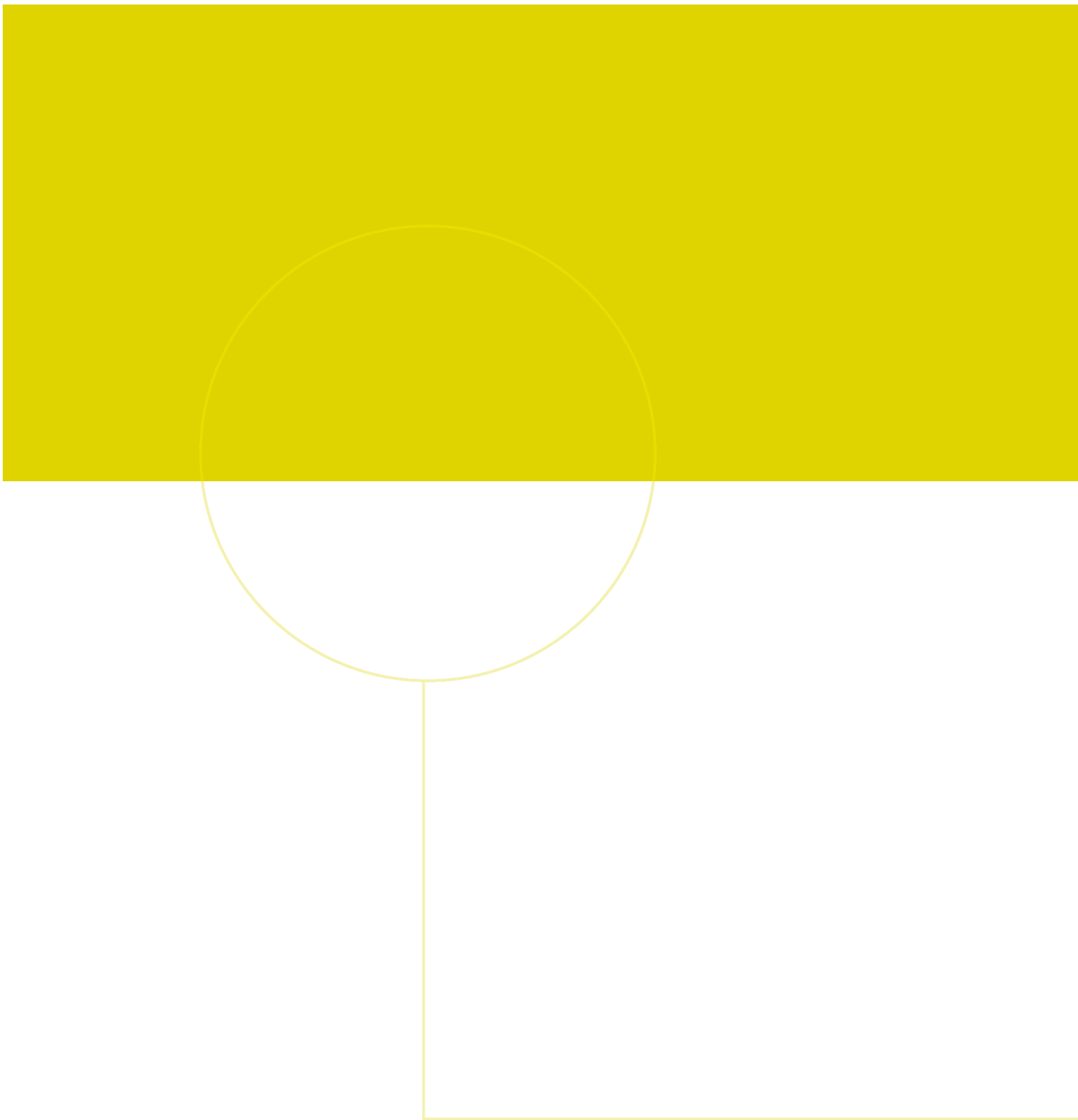
Introduction: When looking at EEG, sleep spindles are visualized as powerful burst of coherent brain activity, usually with a duration between 0.5s and 3s in the sigma frequency band. Sleep spindles is a defining characteristic for N2 sleep and is apparent in all stages of NREM sleep. Specifically fast spindles (12-16Hz) are reported to critically contribute to memory consolidation, where detections of these are highly reliant on manual annotations. Many of the automatic detection algorithms require large and diverse data, which entrust data labeled through visual scoring of sleep experts. This is time-consuming, insufficient and can introduce inter-rater reliability scorer bias.

Methods: To overcome prejudiced annotations used in prediction algorithms, a mathematical-based model was introduced to emulate the sleep expert. The A7 algorithm is based on spindle detection for N2 sleep, where four thresholds need to be exceeded simultaneously to verify a spindle. The calculations are run on artifact-free data to obtain self-reliant spindle annotations, not prone to prejudice or comparison. The output of the A7 is then used as annotations for the DOSED method, a novel deep learning architecture used to predict micro-architecture events. The features of the raw EEG signals are represented by a convolutional neural network, with two modules repeatedly performing localization and classification. The manual workload for the sleep expert would rather involve verification of outputs, where consistency in detection is ensured without human preconception.

Results: The detections methods chosen for the spindle detection have individually performance that surpass individual human experts, where both methods have F1-score above or around 0.70. The method was tested on a limited dataset with eight subjects, to test if the prediction outcome could imitate expert annotations without intervention by sleep expert. The only intervention needed from the sleep expert was validation of a selected percentage of data.

Conclusion: The findings show that there is potential with the first of its kind semiautomatic spindle detection platform, where an independent model annotate data used to predict unknown datasets. More data would be required to get more accurate results.

Support (if any): None



NTNU

Kunnskap for en bedre verden