Bachelor's Thesis Computing Science

Can a smartwatch measure sleep onset?

Walter Levens
s1038123

March 22, 2023

First supervisor/assessor:
Dr. Pieter Koopman

Second assessor:
Dr. Peter Achten

Radboud University

Abstract

Sleep onset is the sleep stage in which you transition from wakefulness to sleep. This sleep stage is interesting for optimising power naps, detecting sleep disorders or performing Targeted Dream Incubation for example. Targeted Dream Incubation is a practice where one's dreams are steered towards a certain topic. Currently, sleep onset can only be accurately detected by medical-grade equipment, which must be operated by trained professionals. If sleep onset can be measured by a smartwatch, people can do these measurements at home, without the need for expensive equipment or personnel.

This thesis explores a setup with which a smartwatch can detect sleep onset. The smartwatch we chose in this thesis is the Fitbit Sense. We used this smartwatch to obtain actigraphy data and heart rate data to determine sleep onset. We conducted an experiment in which participants were asked to take a power nap while wearing the smartwatch. Afterwards, we used two algorithms to convert the obtained data from the accelerometer and heart rate sensor into analysable data about sleep. We compared these results with the subjective experiences of the subjects and at the end, the setup of the experiment was evaluated. The setup used in the thesis is not suitable for measuring sleep onset with a smartwatch. The used algorithms are not accurate enough to detect sleep onset, but we expect that the algorithms are accurate enough once they are adjusted to personal sleeping behaviour.

Contents

Chapter 1

Introduction

Sleep onset is a sleep stage, which is the transition from wakefulness to sleep. It has been researched in a recent paper from Horowitz et al. [1] for example. This paper explores how sleep onset is used to influence our dreams. Horowitz et. al. created a device called *Dormio*, which performs Targeted Dream Incubation (TDI). TDI is a practice in which you try to steer one's dream towards a certain topic. This is done by repeatedly exposing the subject to auditory stimuli during sleep onset. In this sleep stage, up to 70% of people experience dreams [2] while they still process external information [3, 4]. This means that if the subject is exposed to stimuli during sleep onset, there is a chance that their brain incorporates these stimuli into their dream. This can, for example, be useful for reducing nightmares or it can help solve problems in our daily life [5]. As Horowitz et al. [1] noted, TDI is still a relatively new concept in science, so there is much to be explored. This means that research in sleep onset benefits research in TDI.

TDI is not the only reason why sleep onset is an interesting sleep stage. It also helps detect certain sleep disorders. The usefulness of sleep onset is important in daytime sleepiness tests like the Multiple Sleep Latency Test (MSLT) and the Maintenance of Wakefulness Test (MWT) [6]. These tests are used to measure excessive daytime sleepiness in patients. Detecting sleep onset is an important part of both tests because it indicates how long it takes for someone to fall asleep (MSLT) or whether the person is awake (MWT). The results of these tests indicate whether a patient has certain sleep disorders like narcolepsy, idiopathic hypersomnia or sleep deprivation. These tests are important for diagnosing sleep disorders and finding appropriate treatments.

Sleep onset is also interesting for optimising power naps [7]. Research on power naps concludes that the optimal time for a power nap is around 10 to 30 minutes [7, 8]. This length generally reduces some fatigue, without entering deep sleep (Section 2.1). This is beneficial because if you get woken up while in deep sleep, you can suffer from sleep inertia, achieving the opposite of the purpose of power naps. It can, however, be difficult to achieve this optimal sleep time since you cannot determine beforehand at what speed you fall asleep and when you are drifting into deep sleep. If you can measure sleep onset accurately, you can optimize the effectiveness of a power nap by making sure you do not enter deep sleep. This generally improves cognitive performance directly after the nap. [9].

While you enter sleep onset, some physiological changes occur. For example, your heart rate drops, your movement subsides and your brain activity changes [3, 4]. If you accurately measure these physiological changes, you can detect sleep onset. The golden standard for doing this is Polysomnography (PSG, section 2.2.5). PSG is a method that involves complex devices and it requires specific expertise to operate these devices and evaluate the obtained data [10]. Examples of such devices measure brainwaves or eye movement. Because of the complexity, it is mostly done in laboratories and it must be supervised by experts. This is very costly and time-consuming, so it is not accessible to the general public.

Several doors open once the general public has access to a relatively precise method of detecting sleep onset. They can optimize their power naps and perform some kind of TDI by themselves. Moreover, basic sleep disorder detection can be done at home with adapted versions of MSLT and MWT. This can speed up the process of detecting anomalies because the patient does not have to go to a laboratory as often to be investigated.

Wearable devices are useful for measuring sleep onset because almost all smartwatches have sensors that monitor heart rate or movement for example. This means that some of the physiological changes that occur during sleep onset can be detected by a smartwatch. Scott et al. [7] combine the results of 71 papers that try to determine sleep onset with a variety of wearable devices. However, most devices are either medical devices, outdated devices or both. Because of this, it is interesting to see whether a relatively new smartwatch can detect sleep onset accurately.

This thesis first elaborates upon some background information and related work (Chapter 2). After this, it discusses why the Fitbit Sense has been chosen for this research (Chapter 3). This is followed by an explanation of the used algorithms, the program to obtain the data and the procedure of the experiments (Chapter 4). Then, the results of the experiments are shown and interpreted (Chapter 5). In the conclusion, there is a reflection on the executed research and there are suggestions for further research (Chapter 6).

Chapter 2

Background: Sleep

This chapter first gives background information on the terminology used in this thesis in the domain of sleep. Afterwards, it explores important related works to this thesis.

2.1 Sleep cycle and sleep stages

Every time we go to sleep, we enter a sleep cycle. Such a sleep cycle consists of different stages. These stages are classified into three categories: wake, non-REM sleep and REM sleep [11, 12]. REM is an abbreviation for Rapid Eye Movement. It gets its name from the fact that during this sleep stage your eyes are moving rapidly [13].

Non-REM sleep is subdivided into three stages: N1, N2 and N3 [11]. N1 sleep is the first sleep stage that occurs when you are falling asleep. It is also called 'the gateway to sleep' because of that. N1 sleep lasts mostly between 1 and 7 minutes. It is characterised by a change in brainwaves and a decrease in blood pressure, heart rate, core body temperature and muscle tension. [3]. N2 sleep is the sleep stage that occurs after N1 sleep as well as between N3 sleep and REM sleep. It generally lasts between 10 and 25 minutes. In this stage, a sleeper is less easily woken up than in N1 sleep, but they can still process external information. N2 sleep is characterised by a further decrease in heart rate, blood pressure and body temperature as well as some unique patterns in brain waves [4]. Together, N1 sleep and the first occurrence of N2 sleep in a sleep cycle are considered sleep onset as they prepare your body to enter deep sleep. N3 sleep is considered deep sleep and it is generally the third sleep stage that occurs during sleep, after N2 sleep. Characteristics of N3 sleep are a further decrease in blood pressure and heart rate. Besides this, blood flow is directed more to the muscles so that they can restore. As a result, the blood flow to the brain is decreased. Because of this, the sleeper is often disoriented and confused when woken up in this stage [14].

Lastly, there is Rapid Eye Movement sleep or REM sleep. During this sleep stage, the heart rate increases, respiration becomes irregular and the brain is active. Because of these reasons, it is often called 'active sleep'. REM sleep is considered a highly important sleep stage because it aids in emotional processing, memory consolidation and brain development for example. REM sleep is often the last sleep stage in a sleep cycle [13]. After REM sleep, the cycle starts over and you shortly enter N1 sleep again.

This completes the basic background on sleep stages. Many sleep stages are more complex than described above. It is, however, not in the scope of this thesis to go in dept into these sleep stages. This thesis is about detecting sleep onset so we focus mostly on sleep stages N1 and N2.

2.2 Sleep measuring methods

As discussed above, sleep is divided into different stages, each with different physiological characteristics. In this section, we first discuss how a period of sleep is divided into parts to make classification easier. After that, we discuss methods that are used to detect the changes in physiological characteristics of each sleep epoch.

2.2.1 Sleep Epoch

In sleep research, sleep is measured using periods called sleep epochs. A sleep epoch is part of a longer recording of sleep. It depends on the study how long an epoch is, but generally, a duration of 30 seconds is used [11]. The data obtained in one epoch is combined to determine what sleep stage a subject is in for example. In this thesis, a sleep epoch is 30 seconds.

2.2.2 Actigraphy

Actigraphy is a method used to track sleep using motor activities. It is the most used measuring technique in sleep research in combination with wearable devices [7]. Actigraphy measures motor activities with an accelerometer. An accelerometer measures movement in terms of acceleration. The acceleration is measured in m/s^2. Sometimes the letter g is used for acceleration, which is the standard gravitational acceleration of 9,81 m/s^2.

Because actigraphy is a widely used measuring technique, there is a clear view of what it is good at and what it is not good at. Actigraphy is often viewed as relatively accurate when monitoring whole nights of sleep for several days or weeks. However, as Scott et. al show in their paper, the state in which you lay still while being awake is often classified as sleep by actigraphy. This results in an overestimation of Total Sleep Time (TST) and an underestimation of wakefulness. Despite this, we decided to use actigraphy as a way to measure sleep onset.

2.2.3 Blood-oxygen saturation (spO2)

Blood-oxygen saturation (spO2) is a measurement of how saturated your blood is with oxygen. SpO2 is mostly measured using pulse oximetry, where a small clip is placed on the index finger. From the pulse of the bloodstream, the sensor in the clip measures the saturation of oxygen. In smartwatches, spO2 is also measured using the pulse of the bloodstream. Instead of on a finger, a smartwatch measures spO2 on the wrist. This is less accurate than measuring it on the fingers because the blood vessels are further away from the surface of the wrist, compared to blood vessels in the finger [15]. Blood needs to have a certain level of oxygen for organs to function. Normally, spO2 varies between 95% to 100%. This percentage slightly decreases when in sleep because of the decreased respirational activity. Generally, it drops to between 90% and 96% [16]. This is why it is an interesting parameter to measure sleep.

2.2.4 Heart Rate (HR)

As said in section 2.1, heart rate changes while asleep. It drops during sleep onset, which makes it an interesting parameter to detect sleep onset.

There are typically two types of heart rate sensors used in smartwatches: photoplethysmography (PPG) and Electrocardiography (ECG). These two sensors measure heart rate in two different ways: one with reflections of light and the other one with electrical activity.

Photoplethysmography (PPG)

Photoplethysmography (PPG) is the standard heart rate sensor in smartwatches. This technique uses the reflection of light to determine for example the heart rate. The reflection of light is influenced by the volume of the blood flow. The variation in the blood flow changes the reflection intensity, with which PPG determines HR. This is an indirect way of measuring heart rate because it uses the pressure of blood flow instead of the beat of the heart itself. PPG needs a high concentration of blood vessels to take measurements. Because of this, it is generally not that accurate in measuring small changes in heart rate, but it is a good way to monitor heart rate for one or multiple successive days [17].

Electrocardiography (ECG)

Electrocardiography (ECG) measures heart rate based on the electric activity the heart produces each time it contracts. It is a relatively new method to measure heart rate in smartwatches. Medical grade ECG requires patients to place electrodes on the skin near the heart, so it is as close to the source of the electrical activity as possible. With wearables such as smartwatches, this is different, because it measures from the wrist and it does not make use of electrodes. This means that ECG measurements from smartwatches are less accurate than ECG measurements from medical devices. It is, however, more accurate in measuring heart rate than PPG [17].

2.2.5 Polysomnography (PSG)

Today, polysomnography (PSG) is considered to be the golden standard for measuring and monitoring everything involving sleep. It is often used in sleep research and in detecting sleeping disorders in patients. A typical PSG measurement takes into account brain activity, eye movement, muscle response, heart rate, blood-oxygen saturation and respiratory activity. Because of this, PSG is a complex method to measure sleep and it needs experts who calibrate and operate the measuring devices. The results also have to be interpreted by a trained professional or a sophisticated algorithm. The advantage of such a diverse set of parameters is that it accurately detects patterns and abnormalities in a patient's sleeping behaviour [10]. The complexity of PSG makes it an accurate, but inaccessible method to measure and monitor sleep at home. This means that this thesis does not make use of PSG.

2.3 Related work

Scott et al. combine the results of 71 papers about the accuracy of measuring sleep onset with wearable devices [7]. These devices mainly are actigraphy devices, but also devices that measure eye movement or brain waves are included. For this thesis, the findings on actigraphy are most relevant. They found that the results of the different devices and algorithms varied a lot. Some papers reported actigraphy devices to be very accurate, while others found a lot of variation in the performance of their actigraphy devices. This wide variation in results had several reasons, but the most important one was that the results depend heavily on the individual tested. Everyone has different sleeping behaviour, so one algorithm does not do the job. This is why Scott et. al. concluded that the algorithm of actigraphy devices should be adjusted to the individual for the best results. Because of this, we try to find algorithms that automatically adjust to the subject.

An example of a study in which HR is used to identify sleep stages is Fonseca et. al. [18]. In this study, they used PPG to identify sleep stages. Their findings are that PPG is a promising technique to monitor sleep for long-term sleep monitoring. It does, however, need some more research to find out if it can be used as a cheap and more accessible alternative to PSG.

Chapter 3

Selection of the smartwatch

There is an extensive variety of wearable devices capable of measuring sleep accessible to the general public: activity trackers, smartwatches, headbands, rings or earplugs for example. We decided to focus on smartwatches because this kind of device is owned by many people and on most smartwatches, app development is supported. This chapter talks about which smartwatch is chosen and why. Because of a limitation in time and resources, we decided to only consider one smartwatch instead of multiple. Considering multiple watches and comparing their results is something that can be done in future research.

3.1 Smartwatch

Once we decided to execute this research with a smartwatch, we needed to find the most suitable one for us. We determined this by considering the following:

- **The watch should be useful for the general public.** It should be compatible with most smartphones.
- **Sensors.** The watch should have sensors that are useful for detecting sleep. See Chapter 2 for a detailed description of the useful sensors. We need an accelerometer and a heart rate sensor that measures heart rate (HR) and if possible blood-oxygen saturation (spO2)
- **Accessibility to Sensor data.** We need the data from the sensors to apply our algorithms to it.

Table 3.1 contains an overview of smartwatches that were considered, as well as a more elaborate explanation. Note that this list is not exhaustive. Some watches were already disregarded after noticing that one of the conditions was not met. This means that further research on the devices was not done and thus some information in the table is not filled in. Besides, we did not consider every watch that is available on the market. We limited our search scope to brands that are well-known and watches that were generally reviewed positively.

In the table are a few abbreviations. In column 'sensors', 'acc' means accelerometer. In the column 'Verdict', X means not useful and O means useful.

Smartwatch	**Compatible**	**Sensors**	**Sensor access**	**Verdict**
Apple watches	Only iOS	-	-	X
Garmin Vivowatch 4	Android, iOS	PPG, acc, spO2	-	X
Samsung Galaxy Watch 4	Android, iOS	PPG, acc, spO2	Need Privileged Health SDK	X
Samsung Galaxy Watch 5	Android, iOS	PPG, acc, spO2	Need Privileged Health SDK	X
Fitbit Versa 3	Android, iOS	PPG, acc, spO2	Direct access to sensors	X
Fitbit Versa 4	Android, iOS	PPG, acc, spO2	Development not supported	X
Fitbit Sense	Android, iOS	PPG, acc, spO2	Direct access to sensors	O
Fitbit Sense 2	Android, iOS	PPG, acc, spO2	Development not supported	X

Table 3.1: Smartwatches that were considered for this research

To determine whether or not a smartwatch is accessible to the general public, we focused on compatibility. This is important to make sure that we do not include watches that can only be used with a specific brand or version of a smartphone. With this criterion, we quickly exclude Apple watches since they are only compatible with iPhones.

Besides compatibility, an important aspect of the smartwatch is which physical sensors it contains. Physical sensors that are most useful for the research are an accelerometer and a heart rate sensor. Especially the heart rate sensor is interesting because there are two versions of it. After some research, however, we found out that watches that did contain an ECG sensor did not use it to monitor the heart rate passively. It is only used when the user launches a specific application where the heart rate is measured for only 30 seconds. With this application, the user is required to wear the watch on one wrist while also pressing two buttons or sensors on the watch with the index finger and thumb of the other hand. Naturally, this process is not useful if participants need to fall asleep, so we have to use PPG to measure the heart rate.

Besides the accelerometer and the heart rate sensor, we also want the data for Blood-oxygen saturation (spO2). After some searching, we found the following candidates:

- Samsung Galaxy Watch 4
- Samsung Galaxy Watch 5
- Fitbit Sense
- Fitbit Sense 2

Finally, we have to make sure that we can access the data from the sensors. For the accelerometer, this is the acceleration of the watch in m/s^2 and for the heart rate sensor, this is the heart rate in beats per minute (BPM). The calculation of the blood-oxygen saturation is quite difficult. Because of this, we want to use the values for spO2 provided by the API of the watch. After some research, we discovered that both the Samsung watches and the Fitbit watches were not able to provide spO2 data. We cannot use the Samsung watches because we need access to their Privileged Health SDK to obtain spO2 data. At the time of writing, they do not accept new partnerships, so we cannot make use of the SDK. The Fitbit devices need a calibration time of one hour, which is too long for the research we intended to do (see section 4.3). This means that we cannot use spO2 data in our research and thus we do not consider accessibility to spO2 data in our choice for a smartwatch. This is why we only use actigraphy and heart rate to determine sleep onset in this thesis.

We disregarded the Fitbit Sense 2 because it does not support the development of third-party apps. The Fitbit Sense, the Samsung Watch 4 and the Samsung Watch 5 all have extensive documentation, with all the things we need [19,20]. Because of this, there is little difference between the devices for our research. In the end, we decided to pick the Fitbit Sense, because we already looked into working with the Fitbit Sense 2 before we discovered that it does not support the development of third-party apps.

This concludes our research on which smartwatch we choose for this research. The Fitbit Sense passes most requirements we set. So, despite the disadvantage of not being able to measure spO2 in real-time or the heart rate with ECG, the Fitbit Sense is the smartwatch we use for this research.

Chapter 4

Setup of the experiment

This chapter discusses the methods of the research. It first discusses what algorithms are used, followed by the design of the application. After this, the procedure of the experiment is explained.

4.1 Algorithms

For the research, we use the accelerometer and the heart rate sensor to determine sleep onset. To get useful results from these sensors, we need a way to interpret the obtained data. For this we found algorithms to convert the raw data of the sensors into useful data. This section first discusses the algorithm for actigraphy, which uses the accelerometer. After this, it elaborates upon the algorithm for interpreting the heart rate. Finally, it explains how the results of both algorithms are combined to draw useful conclusions.

4.1.1 Actigraphy

The paper of Scott et. al. covers many different papers that use some kind of actigraphy device [7]. We looked for usable actigraphy algorithms between these papers. Many of these papers use a medical device with an algorithm that is provided by the manufacturer of the wearable. These algorithms are specific to the device and are not open source in most cases so we cannot use the algorithms these devices used. A few papers did mention which algorithm they used and after some research, we found the following algorithms:

- Cole Kripke algorithm [21]
- Sadeh algorithm [22]
- Algorithm of Kuo [23]

The algorithms of Cole-Kripke [21] and Sadeh [22] were used in multiple papers. They are viewed as standards in sleep-wake identification using actigraphy and they have been evaluated thoroughly. Both of these algorithms consider the values of several epochs before and after the current epoch. The algorithm of Cole-Kripke has a separate weight for each of the epochs that are considered. The algorithm of Sadeh uses a polynomial with calculations like the mean activity and standard deviation of activities during multiple epochs. Because of this dependence on several epochs after the current epoch, the algorithms are not that useful to measure sleep onset in real-time. Besides, both algorithms use specific medical actigraphy devices that compute an activity score. The activity score is based on how often the intensity of movement is higher than a threshold value within an epoch. It is not clear from the papers or the webpage of the manufacturer of the devices what this threshold is or how it is computed. Because of these two reasons, both of these algorithms cannot be used for this research.

This leaves us with the algorithm of Kuo et al. [23]. This algorithm computes a threshold, instead of using set ones. In addition to this, the algorithm they created is elaborately described and the results of it were promising. It has not been tested and evaluated as thoroughly as the other two. Despite this, we decided to use this algorithm because it fits this research after a few modifications.

The algorithm of Kuo et. al.

The algorithm of Kuo et. al. uses two methods to determine whether the subject is asleep or not. The first method uses the peak-to-peak interval. Peaks in this context are peaks in the acceleration of the device. The peak-to-peak interval method computes the minimal interval between two successive peaks in each epoch. It only considers peaks that exceed a certain threshold. In this case, the threshold is a range around the mean actigraphy values of the epoch that is considered. The values of the range in this research are the same as in the paper: 3.35mg above and below the mean. Here, 1g = 9.81 m/s^2, which is the nominal gravitational acceleration. The calculation of the peak-to-peak interval can be seen in equation 4.1.

$$PP_i = \begin{cases} \min\ (t_{j+1} - t_j), & N \geq 3 \\ 0, & \text{otherwise} \end{cases} \tag{4.1}$$

Where PP$_i$ is the peak-to-peak interval in epoch i. t_j and t_{j+1} are the times of the jth and j+1th peak within epoch i, which are higher than the threshold. N is the number of peaks above the threshold. If there are fewer than 3 peaks in an epoch it is considered noise.

When this is computed, the algorithm determines if epoch i is considered as 'active', based on the peak-to-peak interval. If $\mathrm{PP}_i < 11$ seconds, the current epoch is scored as 'active' (or 1), otherwise it is scored as 'not-active' (or 0). This is not stored in a Boolean because we use this value in the next step for the calculation. See equation 4.2.

$$Activity\_PP_i = \begin{cases} 1, & PP_i < 11 \\ 0, & \text{otherwise} \end{cases} \tag{4.2}$$

Finally, the algorithm determines the Movement Density (MD) per epoch. MD is calculated using the epochs before and after the current epoch. This is called the window size. In the paper, Kuo et al. differentiate between a large, medium and small window size. A large window size is used for people that are known to move a lot in their sleep and a small window size is used for people that don't move as much in their sleep. The medium size is used to determine whether the subjects need a large or small window size. For simplicity, we decided to only use the small window size in our experiment. This is because we only have a limited time to obtain data per participant. A larger window size would slow down the process of determining sleep onset. In the paper, the small window size is set to 15 epochs. This means that it considers the current epoch, the preceding 7 epochs (3,5 minutes) and the following 7 epochs (also 3.5 minutes). This window size is still too big for determining sleep onset in real-time, especially if we need to consider the 7 epochs after the current epoch. This is why we decided to consider 2 window sizes. One window size is the original window size of 15, used in the paper. The second window size will not consider the epochs after the current epoch, so the window size is 8. We use both window sizes so we can compare their results to see if there are differences between them.

To calculate the MD, we compute the sum of $Activity\_PP_i$ and divide that over the window size. If there are less than 7 epochs before the current epoch, that amount of epochs is considered. This holds for both window sizes. For the window size of 15, we also need to check whether there are 7 epochs after the current epoch. If there are fewer than 7, the remaining amount of epochs is considered. See equation 4.3 for the situation with a window size of 8 and equation 4.4 for a window size of 15.

$$MD_i^{PP} = \begin{cases} \left(\sum_{1}^{i} Activity\_PP_j\right)/i, & \text{when } 1 \leq i < w \\ \left(\sum_{i-w}^{i} Activity\_PP_j\right)/w, & \text{otherwise} \end{cases} \tag{4.3}$$

$$MD_i^{PP} = \begin{cases} \left(\sum\limits_{1}^{i+\lfloor w/2 \rfloor} Activity\_PP_j\right)/(w - \lfloor w/2 \rfloor + i), & \text{when } 1 \leq i \leq \lfloor w/2 \rfloor \\ \left(\sum\limits_{i-\lfloor w/2 \rfloor}^{i+\lfloor w/2 \rfloor} Activity\_PP_j\right)/w, & \text{when } \lfloor w/2 \rfloor < i < L - \lfloor w/2 \rfloor \\ \left(\sum\limits_{k-\lfloor w/2 \rfloor}^{L} Activity\_PP_j\right)/(w - \lfloor w/2 \rfloor + L - i), & \text{when } L - \lfloor w/2 \rfloor \leq i \leq L \end{cases} \tag{4.4}$$

Where i is the current epoch, L is the total amount of epochs and w is the window size. This concludes the first method of the algorithm. Now we take a look at the second method: the maximum magnitude.

The maximum magnitude is calculated in a straightforward way. It first checks whether a peak is above the threshold value (3.35mg). If the peak is above the threshold, it compares the value of the peak with the current maximum value. If the new value is higher, it becomes the new maximum value, if it is not higher, the maximum value stays the same. See equation 4.5.

$$Max_i = \begin{cases} \text{max ACC(t)}, & t \in \text{epoch } i \text{ and ACC(t)} > T \\ 0, & \text{otherwise} \end{cases} \tag{4.5}$$

Where Max_i is the maximum magnitude in epoch i, t is the time of the peak and T is the threshold value.

After this, the algorithm determines whether the current epoch is classified as 'active' or 'not-active'. If $\text{Max}_i = 0$, the current epoch is classified as 'not-active' (or 0), otherwise it is classified as 'active'. Again, this value is not stored in a Boolean because it is used in the calculation of the next step. See equation 4.6.

$$Activity\_Max_i = \begin{cases} 1, & Max_i \neq 0 \\ 0, & Max_i = 0 \end{cases} \tag{4.6}$$

After this computation, we also compute the MD of the current epoch, based upon Max_i. Also here, we use the 2 window sizes as discussed in the first part of this algorithm. This calculation is similar to that of the MD of the Peak to Peak interval. See equation 4.7 for the situation with window size 8 and equation 4.8 for window size 15.

$$MD_i^{Max} = \begin{cases} \left(\sum\limits_{1}^{i} Activity\_Max_j\right)/i, & \text{when } 1 \leq i < w \\ \left(\sum\limits_{i-w}^{i} Activity\_Max_j\right)/w, & \text{otherwise} \end{cases} \tag{4.7}$$

$$
MD_i^{Max} = \begin{cases} \left(\sum_{1}^{i+\lfloor w/2 \rfloor} Activity\_Max_j \right) / (w - \lfloor w/2 \rfloor + i), & \text{when } 1 \leq i \leq \lfloor w/2 \rfloor \\ \left(\sum_{i-\lfloor w/2 \rfloor}^{i+\lfloor w/2 \rfloor} Activity\_Max_j \right) / w, & \text{when } \lfloor w/2 \rfloor < i < L - \lfloor w/2 \rfloor \\ \left(\sum_{k-\lfloor w/2 \rfloor}^{L} Activity\_Max_j \right) / (w - \lfloor w/2 \rfloor + L - i), & \text{when } L - \lfloor w/2 \rfloor \leq i \leq L \end{cases} \tag{4.8}
$$

Where i is the current epoch, L is the total amount of epochs and w is the window size. This concludes the second method of this algorithm.

For the final part of the algorithm, we deviate from the algorithm proposed by Kuo et al. In their paper, Kuo et al. [23] make use of four rules to determine whether the subject is asleep or awake. This binary classification is not in line with the consensus that sleep onset is a slow descent into sleep [24]. This is why we approach this a bit differently. In this research, we compute the mean of the two MDs and plot that in a graph. this combined MD is a value between 0 and 1 so we hope it gives us a gradual line if we plot it in a graph.

This concludes the algorithm used to determine sleep onset with actigraphy. The used code of this algorithm can be found on Gitlab [25].

Although we believe that this algorithm suits our research well, some things could be improved. First of all, one of our window sizes is different than proposed in the paper of Kuo et al. This may result in less accurate results because the algorithm is not adjusted to our window size and the window size is not adjusted to the subject. As stated in many papers that Scott et al. [7] reviewed, the algorithm should be adjusted to the subject to obtain the most accurate data. This means that it might be worth looking into how to determine the window size for a short experiment just like ours so that this algorithm can accurately detect sleep onset. On the other hand, you could use long-term data about a person's sleep behaviour to personalise the algorithm. A smartwatch is very helpful in this because many people already use it to monitor their sleep. This means that a smartwatch already has a good insight into the general sleeping behaviour of the user. This information can be used to specialise the algorithm that can detect sleep onset.

Besides this, the algorithm uses a binary classification of sleep, because it is originally designed to only differentiate between sleep and wake. As stated before, this binary classification is not in line with the consensus that sleep onset is a slow descent into sleep [24]. Because of this, we changed the last part of the algorithm, but there are still parts that use binary classification, for example, the calculation of the Activity in an epoch. The algorithm could be changed so that the classification of $Activity\_PP_i$ and $Activity\_Max_i$ is a value between 0 and 1, instead of either 0 or 1. This could improve the accuracy of the MDs. So it might be worth looking into a way to translate the computations of $Activity\_PP_i$ and $Activity\_Max_i$ to a non-binary version.

Finally, a disadvantage of this algorithm is that the result is a Movement Density. Even though this might give a good representation of how much a subject moves, it does not necessarily say something about sleep. It is interesting to look into a way of connecting Movement Density to for example wake probability like Prerau et. al. did in their paper [24].

These suggestions might be interesting to consider in future research to get a more precise result for experiments like this one.

4.1.2 Heart Rate

To measure the heart rate of the participants, we use the Sleep-Onset Period Detection algorithm designed by Jung [26]. We chose this algorithm because it focuses specifically on sleep onset, unlike the algorithm from for example Kuriwara and Watanabe [27]. Besides this, it has an extensive explanation of how the algorithm works. Another advantage is that it does not require a long calibration time to set up the algorithm. The research is aimed at detecting Obstructive Sleep Apnea (OSA), but the algorithm performs well in the 'non-OSA' control group too.

The algorithm itself consists of two parts: it first needs two minutes to calibrate and after that, it determines what epoch is the first to be defined as a 'sleep-onset epoch'.

The calibration phase of the algorithm is used to compute a threshold value. This value is computed over the first 4 epochs (2 minutes) because then the participant is still awake most of the time. This is important because the heart rate during rest is considerably lower than when you are awake. This way we can easily see the difference in heart rate between wakefulness and sleep. This threshold value is used in the second part of the algorithm to determine a sleep-onset epoch. It is calculated as follows:

$$\mathrm{TH} = \mathrm{HR2_{avg}} - 1.96 \times \mathrm{HR2_{SD}} \tag{4.9}$$

Where TH is the threshold value, $\mathrm{HR2}_{avg}$ is the mean heart rate in the first 4 epochs and $\mathrm{HR2}_{SD}$ is the standard deviation of the heart rate in the first 4 epochs.

In the paper, Jung obtained the heart rates through R-R intervals. An R-R interval is the time between two consecutive heartbeats. In our research, however, we get the values for heart rate from the device, we do not compute these values manually. The algorithms Fitbit uses are proprietary, so we cannot say exactly how the values are computed.

This concludes the first phase of the algorithm. The second phase of the algorithm starts in the fifth epoch. From this point on, the algorithm listens for a sleep-onset epoch to determine the sleep onset period. It does this by examining each heart rate sample and comparing that to the threshold

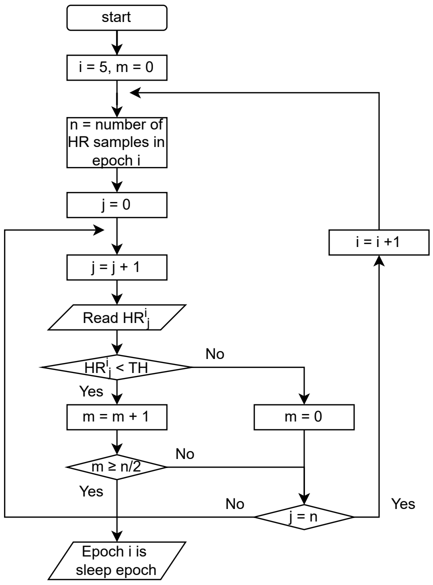


Figure 4.1: Flow diagram of the Sleep-Onset Period Detection algorithm as described by Jung [26]. Explanation of variables: i = the number of the current epoch, j = the number of the current sample, m = current amount of samples below TH in succession, TH = threshold value.

value computed in phase 1. If more than half of the samples in one epoch are successively beneath the threshold value, the epoch is scored as a sleep onset epoch. The sleep onset period is defined as the sleep onset epoch, the two epochs before that and the two epochs after that. See figure 4.1 for the flowchart of this algorithm. We do not have to tweak this algorithm to fit our research. The used code of this algorithm can be found on Gitlab [25].

Even though this algorithm fits our research well, it does have some disadvantages. Just like the actigraphy algorithm, it calculates sleep onset with the use of a threshold. This makes the calculation binary, neglecting the fact that sleep onset is a gradual descent into sleep [24]. For future research, it might be interesting to find a way to translate this binary identification to a gradual scale. This might make the results more accurate.

Besides this, our current setup calculates the threshold based on the first 4 epochs of the recording using a general calculation. Even though this general calculation computes the threshold relative to the heart rate of the subject, it does not incorporate the personal sleeping behaviour of the participant. Personal sleeping behaviour is, however, important to accurately determine sleep onset. It could be incorporated into the algorithm if we have more sleeping data on the subject. As said before, smartwatches are very useful for collecting general sleeping data because they are often worn regularly. Because of this, they have a good indication at which heart rate values the subject is likely asleep.

The last disadvantage is that this algorithm stops when it identifies the first sleep onset epoch. It does not consider the fact that a person can wake up during a measurement and fall asleep again. It could be interesting to also include sleep onsets that occur after the first sleep onset.

4.1.3 Combining the algorithms

Now that we have elaborated upon how we implement the algorithms specific to the sensors, we also need to combine the result of both of them. We do this by plotting the results from the actigraphy algorithm in a graph. When this is done, we indicate in the same graph which epochs are considered as the sleep onset period according to the Heart Rate algorithm. We do this per participant. When combining the results, we hope to find a correlation between them. For example, the sleep onset period often starts when the Movement Density is at a certain value. It could also be the case that the results do not correspond well with each other. In that case, we try to find out what causes this and how it could be improved. To check whether the results reflect the participant's sleep, we use their subjective experiences. See Section 4.3 for more information.

There is a big drawback of combining the results of these two algorithms into one result, namely that the types of results are quite different. It might be hard to find a correlation between these two results since one has a data point for each epoch and the other one just marks five epochs as the sleep onset period. We think, however, that it is not impossible to do this, so we decided to stick with these two algorithms.

4.2 App design/program design

The program we use for this research consists of 4 parts: the smartwatch, the companion app on a mobile phone, a local server on the same mobile phone and the storage of that phone. The local server is needed for storing files in the local storage of the phone because the companion app does not support that. The companion app is designed by Fitbit to be the communication between the smartwatch and other APIs of Fitbit as well as an extra runtime environment. Unfortunately, this does not include writing files to the storage of the mobile phone. The architecture used for this research is based on an architecture made by Peter McLennan [28, 29]. We chose this architecture because it already captures and stores raw actigraphy data. This is exactly what we need for our research. It means that we only have to extend the functionality to also include capturing and storing data from the heart rate sensor. For a schematic representation of the architecture, see figure 4.2.

This flow handles the files with the raw data obtained from the watch, so no algorithm is applied to it yet. This is done on a computer at a later

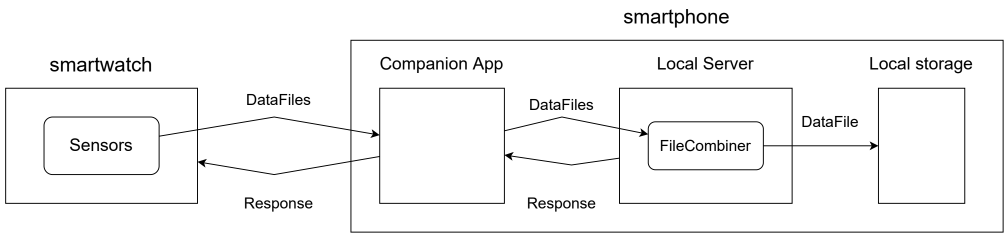


Figure 4.2: Schematic view of the architecture of the program

stage.

Both the smartwatch and the companion app use JavaScript as the programming language, combined with some APIs provided by Fitbit [19]. The local server on the mobile phone is an android app written in Java using Android Studio. The code for the architecture can be found on Gitlab [25].

4.2.1 Smartwatch app

The smartwatch collects the data from the sensors and puts that in a local file. Each file contains data for one epoch (30 seconds) of a sensor. The names of the files are constructed using the sensor name (in our case 'acc' and 'HR') and the file number. We do this to make sure files of different sensors are not mixed and to make sure the proper order is kept.

The accelerometer collects data at a frequency of 10 Hz and does that with an accuracy of 0.002 m/s^2. The heart rate sensor does this at a frequency of 1 Hz and rounds it to the nearest integer. The frequency of the accelerometer is higher because acceleration typically varies more in a second than heart rate. Measuring multiple heart rates per second does not add to the accuracy of the data. We chose a frequency of 1 Hz because this is the highest possible frequency for the Fitbit Sense. It might, however, not be the most ideal frequency because measuring HR at this rate drains the battery. For this thesis, we decided to stick to this high frequency of HR so that we have many data points per epoch.

After each epoch, the current file is closed and a new one is opened. Once a file is closed, it is sent to the companion app using the file transfer API provided by Fitbit [30]. This happens file by file to make sure every file is properly received by the companion app. Once a file is received by the companion app, the smartwatch gets a response message back from the companion app. If this message indicates that the file has been properly processed, the smartwatch sends the next file to the companion app. If the file was not correctly received, the smartwatch tries to resend it. This happens if the file gets corrupted while sending or when the server is not running for example. Only after a new recording is started, the files of the previous recording are deleted. This ensures that no data is lost when something goes wrong in the process of transferring.

If the user wants to start a recording, they press the "start recording" button on the smartwatch. From this point on, the sensors start collecting data. After every epoch, the files are sent to the companion app. Each time, a text field on the smartwatch is updated to indicate which file has been sent to the companion app most recently. If the user wants to stop the recording, they press the "stop recording" button on the device. This stops the recording and sends the last files with data to the companion app. From this point, the user can start a new recording or resend all files in case something went wrong.

4.2.2 Smartphone: Companion app

The companion app serves as a forwarder and communicator between the smartwatch and the local server. Once it receives a file from the watch, the companion app forwards it to the local server. When the file has been sent, the companion app receives a reply from the server about whether the transfer was successful. The companion app forwards this message to the watch. The watch replies to these status reports as described above. The companion app does not store any of the files.

4.2.3 Smartphone: Local server

The local server is used to store the files received from the companion app. Once it receives a file from the companion app, it checks whether the data is not corrupted or if the file has the correct length for example. If it does not pass these tests, it sends a corresponding status message to the companion app and the file is discarded. If the checks are passed, the server stores the files temporarily. Once all files are received, the user can press a button to download the data. There is no check present in the local server that monitors whether the last file is received. In this setup, the user is responsible to check whether all files are received. Once this button is pressed, all files of a sensor are appended in one file. For this part, the names of the files are important to make sure the sensors do not get mixed up and to make sure that the files are appended in the right order. The file with all data of one sensor is then stored in the storage of the mobile phone. This process is repeated for all the sensors that are used. In our case, this means that we are left with two files: a file for the accelerometer data and a file for heart rate data.

4.2.4 Processing of obtained data

To determine sleep onset, we need to convert the obtained data into something useful. For this, we use the algorithms discussed in section 4.1. We create a script in which we implement the algorithms in the programming language C. The two files we obtain from the server are the inputs for this script. The result of this script is a file containing the processed data. These results are plotted in a graph using Libre office, so they can be interpreted (See Chapter 5 and appendix B for more details).

4.2.5 Evaluation of the setup

With the current setup, we can try different algorithms and versions of algorithms on the obtained data, which is very useful for research purposes. We can do this because we only evaluate the data after the experiment and not during the experiment. The drawback of this setup is, however, that we cannot say something about someone's sleep in real-time. If we want to create an app to improve power naps, for example, we need the data close to real-time. This is, however, something for future research.

This setup does have some disadvantages. One of them is that this setup is inconvenient since it consists of four different parts. With the transfer of data between each component, there is a chance that the data gets lost or corrupted. Even though basic checks for this are implemented, it does not guarantee that all data is properly transferred.

Besides this, the current setup is not efficient in terms of battery power, because it needs to transfer quite a large amount of data. This drains the battery of the smartwatch and the mobile phone. This should be looked into if the setup is used in future research.

4.3 Procedure

The experiment is held in a room in a study space at Radboud University. This room is secluded from the main study space so there is little distraction from the outside. The room contains a comfortable chair in which the participants try to fall asleep, see picture 4.3. It also has an adjacent room for the experimenter to be close to monitor the participant during the experiment.

To determine whether the Fitbit Sense can accurately measure sleep onset, we execute the following experiment: Participants are asked to come to the location described above. They are instructed to wear the Fitbit Sense on their left wrist and hold an object in their right hand. This object is either a water bottle or a pencil bag. Both the smartwatch and the object are given to them at the location. After the participant received the watch and the object, they try to take a power nap in a comfortable chair. They are instructed to hold the object in such a way, that it falls on the floor if their muscle tension decreases. This decrease in muscle tension happens as the participants descend into deep sleep. Eventually, the muscle tension is low enough for the object to slip out of their hand and fall on the floor. The sound of the falling object should wake them. When muscle tension reaches this point, it is often an indication that the participants are drifting toward

Figure 4.3: The comfortable chair used in the experiment. It is inside the separate room in which the participants are instructed to sleep.

deep sleep. We don't want people to enter deep sleep because that is not in the scope of our research. Besides, it most likely wakes them up feeling more tired than before they went to sleep, which is the opposite effect of a power nap. If participants are not woken up in this way within 30 minutes, an alarm goes off. The experiment ends when the participant is woken up or when the alarm goes off. At this point, the participants are instructed to report whether they thought they were asleep or not. We use this subjective judgement to determine whether the smartwatch can detect sleep onset. The specific instructions given to the participants can be found in Appendix A.

A disadvantage of this experiment is that we make use of the subjective experiences of the subjects. Although this gives a good indication of how well people slept, we cannot know for example at what exact time they started to doze off. A comparison with Polysomnography (section 2.2.5) or medical-grade Electrocardiography (section 2.2.4) would most likely give a more precise insight into this. We tried to arrange this at the start of this thesis, but unfortunately, the laboratories of the university did not have room for our experiments. We also did not have time or expertise to measure with these methods ourselves. This is something that future research could look into.

Chapter 5

Results

This chapter presents the results obtained from our experiments. We first discuss the sample population of the experiment. After that, we elaborate on how the algorithms performed and what the subjective experiences of the participants are. After this, we compare the results of the algorithms to the subjective judgement of the participants and lastly, we discuss the implication of the results of the experiments.

5.1 Sample population

In total, we conducted seven experiments. From these, four turned out to be not useful because of two reasons: in one experiment, the heart rate was not recorded for about half of the time. In the other three, the data was not stored properly because of a bug in our program, together with the fact that the companion app crashed in these experiments. We had to fix this in between experiments so our program would store the data properly. In the end, we are left with a sample population of three people. All of them are male students from the Radboud university, with ages ranging between 21 and 23.

Because of this small population size, we cannot identify trends or outliers in the data. Individual differences have a big effect on the average result, so our main goal is not to evaluate the results we got from the experiments. Instead, we focus on evaluating the experiment itself by looking at the advantages and disadvantages of our setup and unexpected outcomes for example.

5.2 Actigraphy

When running the actigraphy algorithm over the obtained data, we quickly discovered that the threshold used by Kuo et. al. was too low [23]. Their threshold of 3,35 mg translates to roughly 0.0386 m/s^2. With this value, the

Movement Densities of all subjects would not drop below 1. This is probably because the accelerometer in our device is less accurate than the medical-grade device used by Kuo et. al, resulting in bigger variations between data points. Using this threshold makes the computed data of the accelerometer useless, so we decided to change the threshold value. We can do this because Movement Densities are relative values. As long as we use the same threshold for every data set, we can compare our results and say something about them. When we tested the algorithm with different thresholds, we quickly saw that it is difficult to determine the right value for the threshold. As mentioned before, everyone has different sleeping behaviour, so a single threshold value does not work for everyone. In the end, we decided to test the three data sets with two thresholds. This way we can see what effect the difference in threshold has on the results. The two thresholds we chose are 0.1 m/s^2 and 0.15 m/s^2. We chose these values because they roughly are the borders of what thresholds are usable with our results.

As discussed in section 4.1.1, we used 2 window sizes. In general, these window sizes show a similar graph as can be seen in appendix B. The main difference we notice is that using a bigger window size reduces the variation of the graphs. Incidental peaks and drops are flattened, so we see a more gradual line in general. It is difficult to say which window size is the better one, because of the limited amount of data we have.

Both problems mentioned above might be solved when the smartwatch is worn for a longer time than just during the experiment. If the device is worn for multiple successive days and nights, it can for example monitor sleeping behaviour. This way you know when a person moves a lot in their sleep or not. With this information, you should be able to adjust the threshold value and window size to fit the user. This most likely increases the accuracy of the algorithm.

In section 5.5, we compare the results of the actigraphy algorithm to that of the HR algorithm and the subjective experiences. The visual representations of the results of both thresholds can be found per subject in appendix B, as well as a more detailed analysis of the data per subject.

5.3 Heart Rate

When applying the HR algorithm to the obtained data, we noticed that in subjects 2 and 3 the first epoch after the calibration period was immediately identified as the sleep onset epoch. We do not expect this, because it normally takes more than two minutes for people to fall asleep, especially when they are in an unfamiliar environment. This unexpected behaviour is caused by the fact that the mean heart rate of the first four epochs is quite a lot higher than the mean heart rate in epoch 5. For example, in subject 2 this is around 108 bpm and 98 bpm respectively. The threshold is determined to be

99 bpm because of the high heart rate of the first 4 epochs. This means that the algorithm is not necessarily wrong, but it seems like it is not designed for drops in heart rate as we see in subjects 2 and 3. Because of this, we tried to compute the threshold differently to see what kind of effect this has on the results. We increased the multiplier in the computation of the threshold from 1.96 to 2.5 (see section 4.1.2). In both cases, this change did not have a big effect, indicating that the threshold might still be too high. We decided, however, not to change the computation of the threshold more than this, because it is hard to say what threshold would give the proper results. If we had HR data from multiple days, we could have adjusted the algorithm to the subjects. With the little data we have, we can only change the HR data so that it fits the actigraphy data without argumentation, which does not give credible results. Besides changing the threshold value to account for this large drop in heart rate, we could also have changed the calibration window. If we would extend this from the first 4 epochs to the first 6 for example, the threshold would change and it might have a positive effect on the results. We decided not to change the algorithm in this way, because of the same reason mentioned above.

We decided to also include the average HR per epoch in the graph to get more insight into the results of the experiments. In all three subjects, this shows a line that gradually descends during the experiment. In none of the graphs there is a sudden drop in HR, which makes selecting 5 epochs as sleep onset period look arbitrary.

Just like with the actigraphy algorithm, the situation might be different when the smartwatch is worn for a longer period of time, instead of just during the experiment. In such a case, the threshold could be calculated based on the heart rate of several days, not on the first 2 minutes of a single measurement. With this, the threshold can be adjusted based on the sleeping behaviour of the user. This could for example handle a big drop in HR better than our current algorithm. Using the HR data of several days would most likely make the algorithm more accurate because it is personalised.

Because of the unexpected results of subjects 2 and 3, we cannot say much about the performance of the algorithm itself. We can say, however, that using a personalised computation of the threshold most likely benefits the accuracy of the algorithm. If the threshold is personalised, the algorithm should be able to handle personal differences better.

In section 5.5, we compare the results of the HR algorithm to that of the actigraphy algorithm and the subjective experiences. The visual representations of the results of both thresholds can be found per subject in appendix B, as well as a more detailed analysis of the data per subject.

5.4 Subjective judgements

When the experiment was finished, the participants were asked about their experience. All three subjects reported not having slept, or at least not very deeply. All of them did, however, experience one or more moments where they nearly fell asleep. They think this happened because they were aroused by the alarm or because of another reason that they could not pinpoint. We expect these moments of nearly falling asleep to be marked as sleep onset in the data. None of the experiments stopped because the subjects were woken up by the falling object, they all ended because of the alarm. Even the experiments from which the data was not useful, did not end because of the falling object. This might be an indication that the object did not slip out of the subject's hand easily enough. This means that even when the subjects were drifting towards deep sleep, the object did not fall to the ground. It could also mean that a timer of 30 minutes is too short since the participants need some time to get comfortable, especially in an unknown room during an experiment. A lack of comfort might also be the reason why some participants did not fall asleep during the experiment.

Furthermore, a participant indicated that they weren't able to fall asleep because they heard noises from other rooms. The chair used in the experiment also was not optimal for taking a power nap because you have to sit up relatively straight. This can also influence one's ability to sleep.

5.5 Comparison

If we take the results together, we barely see any connection between the results of the actigraphy algorithm, the HR algorithm, the average HR per epoch and the subjective experiences of the subjects. Some of the results seem to point to the same conclusion in some cases, but once the threshold is changed, the connection often disappears. Because of this, it is not clear whether the occasional agreement between results is a coincidence or not.

In some cases, however, the actigraphy algorithm seems to be able to detect moments where the subjects were aroused. This can be explained by the sudden movement that could go paired with this, combined with the movements of getting comfortable to fall asleep again. The results of the HR algorithm are hard to interpret because of the results we got with subjects 2 and 3.

5.6 Discussion

We hoped that a combination of parameters would give us an indication of when the participants were in sleep onset, just like PSG does this with multiple parameters. But the algorithms that we used are not in line with the subjective experiences of our subjects. This can have several causes like inaccurate measurements or subjective experiences not being representative of what happened. We believe that the most important cause of the disagreement between our results is the lack of personalisation of the algorithms. The short amount of time that the subjects wore the watch does not incorporate personal sleeping behaviour. This is evident if we compare the results from the different subjects with each other. In both algorithms, we see that a threshold is too low for one subject, while it is too high for another subject. Because we do not have information about the personal sleeping behaviour of our subjects, it is hard to determine the right threshold for each participant. Besides this, the limited amount of data also makes it hard to determine whether resemblances in results are coincidences or not.

The results we obtained from this experiment seem to indicate that the algorithms are not personalised enough. Because of this, we think that the setup used in this thesis does not suffice. Some aspects need to be critically reviewed and improved so that this setup is usable. These aspects could for example be the personalisation of the algorithms, how the data is obtained or how the experiment is executed. It is, however, not in the scope of this thesis to improve this setup because of a limitation in time. Future research could look into improving the used methods so that the results of the experiments are more accurate and more usable.

Chapter 6

Conclusions

6.1 Conclusion

Sleep onset is the first stage of a sleep cycle and it is a gradual descent into deeper sleep. Sleep onset is an interesting stage for research because it can be used to optimise power naps, perform Targeted Dream Incubation and detect sleeping disorders for example. Sleep onset can be detected because of physiological changes in the body. Some of these changes can be measured with sensors present in most smartwatches.

In this thesis, we first explored which smartwatch is most useful for detecting sleep onset. We chose the Fitbit Sense and we used its accelerometer and heart rate sensor. We created a setup with which the smartwatch could send the data of these sensors to a phone, so it could be analysed later. We found the algorithm of Kuo to interpret the data of the accelerometer and we used the algorithm of Jung to interpret the data of the heart rate sensor [23, 26].

We executed experiments in which subjects were asked to take a power nap. They wore the Fitbit Sense during this so we could measure the heart rate and movement of the wrist. The subjects also reported whether they fell asleep or not. In the end, only 3 experiments resulted in usable data. Because of this, we did not try to find trends or outliers and draw conclusions from them. Instead, we evaluated the experiment as a whole and identified which things went well and which things can be improved upon.

With the setup we used in this research, we were able to collect data about the heart rate and movement of the subjects during a power nap. We were able to apply the algorithms to the obtained data, however, the results of the algorithms differ a lot per subject. This is likely because both algorithms are too general. Everyone has a different sleeping behaviour so it is difficult to tell which thresholds fit the subjects. We believe, however, that this setup could be a useful starting point for future work if some elements are reviewed and improved.

6.2 Future work

This thesis is about an exploratory experiment on detecting sleep onset with a smartwatch. Throughout this thesis, we already discussed which aspects could be improved upon per section. This section elaborates more upon general points of interest for future research.

In the current research, we only used two sensors to determine sleep onset. When using two sensors, the general results quickly become inaccurate when one sensor is inaccurate. If we would remove the inaccurate sensor, the results will be influenced a lot and we only have one parameter left. This is not desirable. Adding other parameters like spO2 reduces the influence of a single sensor. This can be looked into in future work.

Besides this, the experiments conducted in this thesis require the subjects to wear a smartwatch that they normally do not wear. This cannot incorporate personal sleeping behaviour. Personal sleeping behaviour is, however, important in sleep research, because you have a better view of which physiological changes indicate sleep. Many people already use a smartwatch to monitor their sleep, so it should have a good insight into their sleeping behaviour. With this data, you can personalise the algorithms that detect sleep onset, which should make them more accurate. This is why future research could look into the possibility of using long-term sleeping information obtained with smartwatches to improve the accuracy of the algorithms.

Detecting sleep onset in real-time is crucial for an app that improves the efficiency of a powernap or an app that can perform a version of TDI or basic sleep disorder detection. Because of this, future research could also look into using the smartwatch and algorithms to accurately detect sleep onset in real-time.

Lastly, future research could look into incorporating objective measurements, instead of subjective judgements from the subjects to validate the results from the algorithms.

This concludes our thesis about detecting sleep onset with a smartwatch. We created a starting point on which future research can build. As discussed above and throughout the thesis, there is still quite some room for improvement. Even though we did not find evidence that with our setup, a smartwatch can detect sleep onset, we believe our research is still useful for future research.

Bibliography

[1] Adam Haar Horowitz, Tony J. Cunningham, Pattie Maes, and Robert Stickgold. Dormio: A targeted dream incubation device. *Consciousness and Cognition*, 83:102938, 2020.

[2] Hypnagogia: How the state between wakefulness and sleep works.

[3] Rose MacDowell. N1: Non rapid-eye movement sleep, stage 1. `https://sleepopolis.com/education/n1-sleep/`. Online, accessed 9-Dec-2022.

[4] Rose MacDowell. N2 sleep: K complexes and sleep spindles. `https://sleepopolis.com/education/n2-sleep/`. Online, accessed 9-Dec-2022.

[5] Elena Sanz. Sleep incubation. `https://exploringyourmind.com/dream-incubation-dreaming-with-a-purpose/`. Online, accessed 11-Jan-2023.

[6] Shannon S. Sullivan and Clete A. Kushida. Multiple sleep latency test and maintenance of wakefulness test. *Chest*, 134(4):854–861, 2008.

[7] Hannah Scott, Leon Lack, and Nicole Lovato. A systematic review of the accuracy of sleep wearable devices for estimating sleep onset. *Sleep Medicine Reviews*, 49:101227, 2020.

[8] Rajiv Dhand and Harjyot Sohal. Good sleep, bad sleep! the role of daytime naps in healthy adults. *Current opinion in pulmonary medicine*, 12(6):379–82, 2006.

[9] Frédéric Dutheil, Benjamin Danini, Reza Bagheri, Maria Livia Fantini, Bruno Pereira, Farès Moustafa, Marion Trousselard, and Valentin Navel. Effects of a short daytime nap on the cognitive performance: A systematic review and meta-analysis. *International Journal of Environmental Research and Public Health*, 18(19), 2021.

[10] Jessica Vensel Rundo and Ralph Downey. Chapter 25 - polysomnography. In Kerry H. Levin and Patrick Chauvel, editors, *Clinical Neu-*

rophysiology: Basis and Technical Aspects, volume 160 of *Handbook of Clinical Neurology*, pages 381–392. Elsevier, 2019.

[11] Edwin M Valladares Terese C Hammond Andres A Gonzalez, Kaveh Rezvan. Sleep stage scoring. `https://emedicine.medscape.com/article/1188142-overview?reg=1#a1`. Online, accessed 20-Dec-2022.

[12] Stages of sleep: What happens in a sleep cycle. `https://www.sleepfoundation.org/stages-of-sleep`. Online, accessed 23-Dec-2022.

[13] Jay Summer. What is rem sleep and how much do you need? `https://www.sleepfoundation.org/stages-of-sleep/rem-sleep`. Online, accessed 9-Dec-2022.

[14] Rose MacDowell. N3: Slow wave sleep. `https://sleepopolis.com/education/n3-slow-wave-sleep/`. Online, accessed 9-Dec-2022.

[15] Blood oxygen level. `https://my.clevelandclinic.org/health/diagnostics/22447-blood-oxygen-level`. Online, accessed 9-Dec-2022.

[16] Robert E Gries and Lee Brooks. Normal oxyhemoglobin saturation during sleep* how low does it go?, 1996.

[17] Ecg vs ppg for heart rate monitoring: Which is best?, Jan 2015.

[18] Pedro Fonseca, Tim Weysen, Maaike S. Goelema, Els I.S. Møst, Mustafa Radha, Charlotte Lunsingh Scheurleer, Leonie Van Den Heuvel, and Ronald M. Aarts. Validation of photoplethysmography-based sleep staging compared with polysomnography in healthy middle-aged adults. *Sleep*, 40, 7 2017.

[19] Fitbit development: Reference. `https://dev.fitbit.com/build/reference/`. Online, accessed 4-Dec-2022.

[20] Get started with wear os. `https://developer.android.com/training/wearables`. Online, accessed 22-Dec-2022.

[21] Roger J Cole, Daniel F Kripke, Twilliam Gruen, Daniel J Mullaney, and J Christian Gillin. Automatic sleep/wake identification from wrist activity, 1992.

[22] A Vi Sadeh, Katherine M Sharkey, Mary A Carskadon, and E P Bradley. Activity-based sleep-wake identification: An empirical test of methodological issues, 1994.

[23] Chih En Kuo, Yi Che Liu, Da Wei Chang, Chung Ping Young, Fu Zen Shaw, and Sheng Fu Liang. Development and evaluation of a wearable device for sleep quality assessment. *IEEE Transactions on Biomedical Engineering*, 64:1547–1557, 7 2017.

[24] Michael J. Prerau, Katie E. Hartnack, Gabriel Obregon-Henao, Aaron Sampson, Margaret Merlino, Karen Gannon, Matt T. Bianchi, Jeffrey M. Ellenbogen, and Patrick L. Purdon. Tracking the sleep onset process: An empirical model of behavioral and physiological dynamics. *PLoS Computational Biology*, 10, 10 2014.

[25] Walter Levens. Thesis sleep onset. `https://gitlab.com/SnevelRetlaw/thesis-sleep-onset`. Online, accessed 4-Jan-2023.

[26] Da Woon Jung, Su Hwan Hwang, Yu Jin Lee, Do-Un Jeong, and Kwang Suk Park. Apnea–hypopnea index prediction using electrocardiogram acquired during the sleep-onset period. *IEEE Transactions on Biomedical Engineering*, 64(2):295–301, 2017.

[27] Yosuke Kurihara and Kajiro Watanabe. Sleep-stage decision algorithm by using heartbeat and body-movement signals. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, 42:1450–1459, 2012.

[28] Peter McLennan. Gondwanasoft/fitbit-accel-fetcher. `https://github.com/gondwanasoft/fitbit-accel-fetcher`. Online, accessed 4-Dec-2022.

[29] Peter McLennan. Gondwanasoft/android-fitbit-fetcher. `https://github.com/gondwanasoft/android-fitbit-fetcher`. Online, accessed 4-Dec-2022.

[30] Fitbit development: File-transfer api. `https://dev.fitbit.com/build/reference/device-api/file-transfer/`. Online, accessed 4-Dec-2022.

Appendix A

Instructions experiment

Thank you for participating in this experiment! In a minute you will be trying to take a power nap using the technique from Thomas Edison himself! To make this experiment run smoothly, please adhere to the following instructions.

- Wear the smartwatch on your left wrist. Make sure it is comfortable, but not too loose since that could influence the data collection process
- While attempting to take a power nap, hold the object that is given to you in your right hand. You should hold it in a way so that it falls on the ground when your muscle tension decreases. This happens slowly when you descend into sleep. The falling object should ensure that you are woken up before you enter deep sleep (meaning you don't wake up even more tired than before you went to sleep).
- Take place in the chair and make yourself comfortable. Please keep in mind that you hold the object in the way described above.
- Set a timer of 30 minutes on the smartwatch. This ensures that you do not sleep for too long if you fall asleep. It also makes sure that you don't spend too long trying to fall asleep if it turns out to be hard.
- Once you are comfortable, launch the 'sleep onset' app on the smartwatch and press the start button. Once you do, you are free to fall asleep.
- If you are woken up by the falling object make sure to hit the 'stop recording' button on the side of the watch as soon as possible to end the data collection. Notify the experimenter that you were woken up.
- You have a maximum of 30 minutes to take a power nap, so when you exceed this time, the alarm will ring. Please also press the 'stop recording' button in this case.

- Once you wake up after 30 minutes or because of the falling object, the experiment is over. Don't try to fall asleep again.
- When the experiment is over, reflect on whether you think you fell asleep. It does not matter whether you succeeded in falling asleep or not, in both cases the data can be interesting for us.
- Report your experience to the experimenter and hand over the watch and the object.
- Finally: Try to relax; Don't try too hard to fall asleep. Data of someone who does not fall asleep is as valuable as data of someone who did fall asleep.

Appendix B

Results of experiments

This appendix contains a detailed description of the results of each participant as well as graphical representations of them. It explains how to interpret the graphs and then discusses each subject's results separately.

Each subject has 4 graphs because we used two versions of both the actigraphy algorithm and the HR algorithm.

The x-axis of each graph represents the number of the epoch and the left y-axis is the Movement Density. The blue line indicates the Movement Density where the window only considers epochs before the current epoch and the current epoch itself. The red line indicates the Movement Density where the window also considers epochs after the current epoch, which is the original algorithm. The right y-axis represents the heart rate in beats per minute of the subjects. This is illustrated with the yellow line. The orange bars represent the five epochs that are classified as the sleep onset period by the HR algorithm.

The title of each graph indicates which subject it belongs to and which version of the algorithm has been applied to the data. The following abbreviations have been used:

- *acc TH: 0.1*: The threshold of the actigraphy algorithm is set to 0.1 m/s^2.
- *acc TH: 0.15*: The threshold of the actigraphy algorithm is set to $0.15 m/s^2$.
- *HR mult. 1.96*: The threshold of the HR algorithm is calculated using the original multiplier of 1.96.
- *HR mult. 2.5*: The threshold of the HR algorithm is calculated using the adjusted multiplier of 2.5.

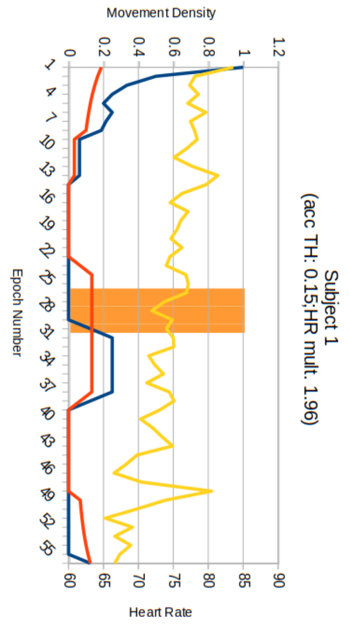
B.1 Subject 1

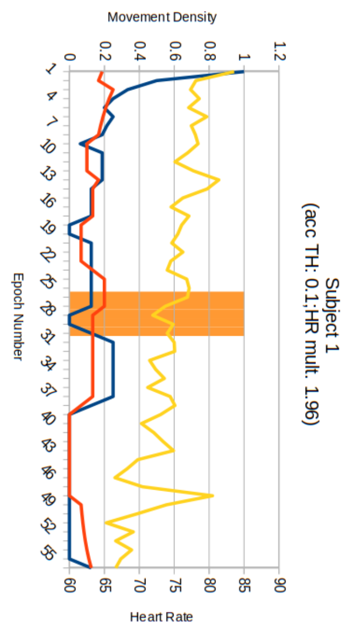
The graphs of subject 1 can be seen on the next page. They seem to indicate that the threshold of the actigraphy algorithm is quite high because in both cases the Movement Density barely exceeds 0.2. The threshold of the HR algorithm seems to be correct because the sleep onset period is somewhere in the middle of the experiment.

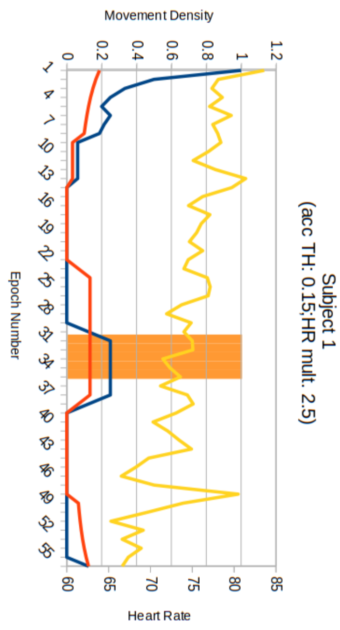
This subject indicated that he was aroused somewhere in the middle of the experiment. We can see an increase in Movement Density in the middle at the threshold of 0.15, with both window sizes. The results of the HR algorithm with the original threshold indicate the sleep onset epoch to be just before the increase in movement density in the graph of the original window size (blue line). This could indicate that this is the moment just before the subject was aroused. When we take the result of the HR algorithm with the altered threshold, this connection is gone.

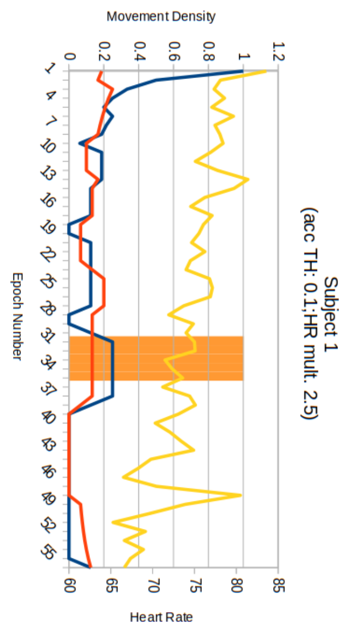
When we take the actigraphy threshold to be 0.1, we do not see an indication of when the subject might have woken up.

The yellow line for HR shows a downward trend during the experiment. It does not show a sudden drop in heart rate at the period classified as sleep onset period by the HR algorithm. This makes the result of the HR algorithm look arbitrary. Besides, there does not seem to be a connection to the results of the actigraphy algorithm, because the heart rate does not increase with an increase in Movement Density.


Subject 1
(acc TH: 0.15;HR mult. 1.96)
Movement Density
Epoch Number
Heart Rate


Subject 1
(acc TH: 0.1;HR mult. 1.96)
Movement Density
Epoch Number
Heart Rate

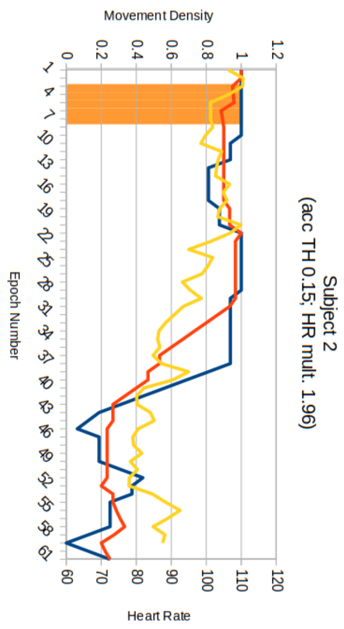

Subject 1
(acc TH: 0.15;HR mult. 2.5)
Movement Density
Epoch Number
Heart Rate

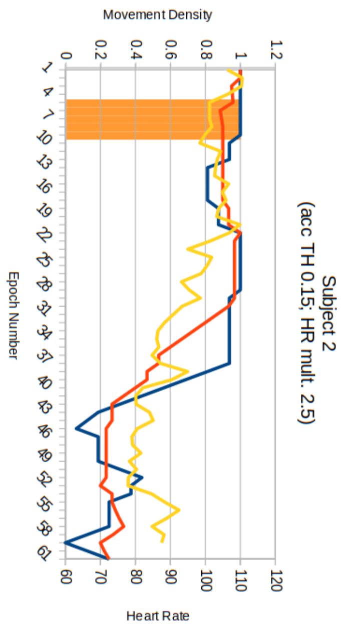

Subject 1
(acc TH: 0.1;HR mult. 2.5)
Movement Density
Epoch Number
Heart Rate

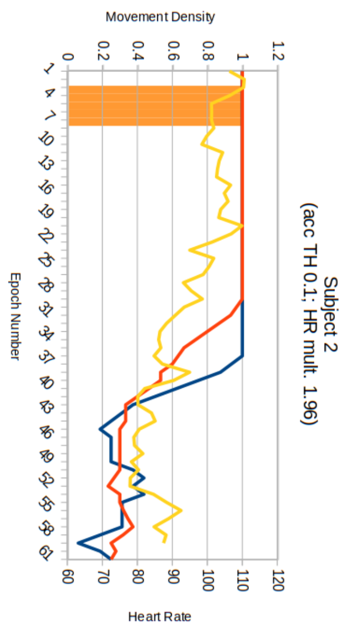
B.2 Subject 2

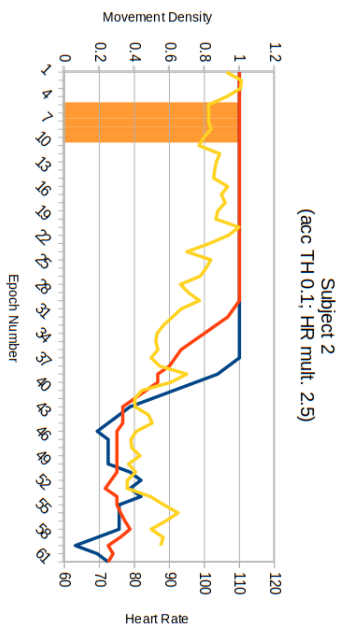
The graphs of subject 2 can be found on the next page. In contrast to the graphs of subject 1, the Movement Density of this subject is very high for the first half of the experiment. This might indicate that the threshold for this subject might be too low. Besides this, the results from the HR algorithm are not helpful because they indicate that the sleep onset period begins at the fifth epoch for the original algorithm or at the seventh epoch for the adjusted algorithm. This is most likely too early because the subjects need some time to get comfortable, especially in our experiment. Because of this, we do not consider the results of the HR algorithm.

Subject 2 says that he felt like the last 5 to 10 minutes went by quickly. This corresponds to the last 10 to 20 epochs. In both graphs of the actigraphy algorithm, we see a quick decline in Movement Density somewhere in the second half of the experiment. This could indicate that the subject fell asleep around that time. This is also supported by the raw Heart Rate data because that shows a decline towards the middle of the experiment. The drop in Movement Densities might also indicate that the smartwatch was first in a position where it detects relatively many movements and later it was moved to a position where it detects them less. Besides this, there are no similarities between the subjective experiences and the results of the actigraphy algorithm.


Subject 2
(acc TH 0.15; HR mult. 1.96)
Movement Density
Heart Rate
Epoch Number


Subject 2
(acc TH 0.15; HR mult. 2.5)
Movement Density
Heart Rate
Epoch Number

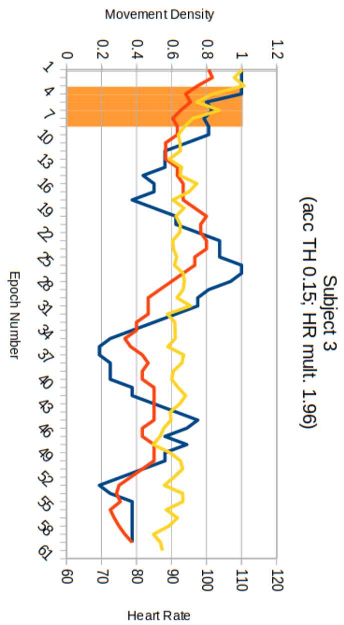

Subject 2
(acc TH 0.1; HR mult. 1.96)
Movement Density
Heart Rate
Epoch Number

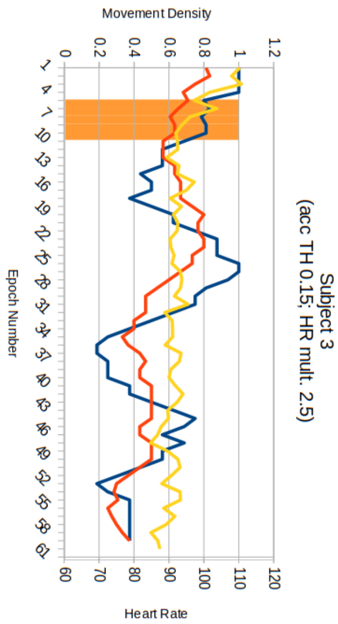

Subject 2
(acc TH 0.1; HR mult. 2.5)
Movement Density
Heart Rate
Epoch Number

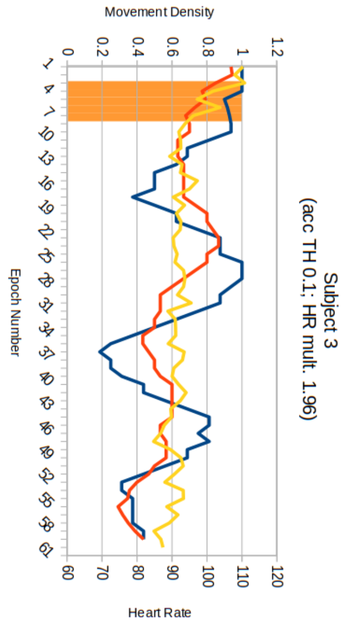
B.3 Subject 3

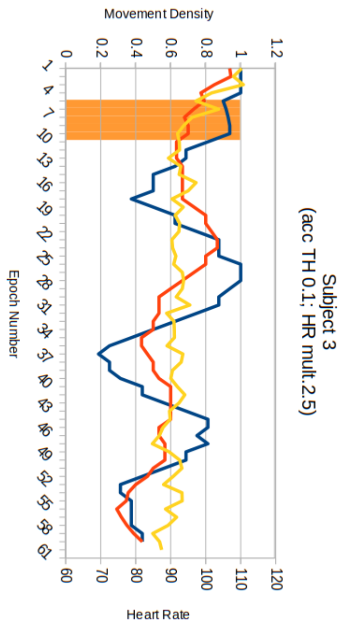
The graphs of subject 3 can be found on the next page. The threshold for the actigraphy algorithm fits this subject well. The values of the Movement Density are mostly between 0.2 and 0.8 in all graphs. We do not look at the results of the HR algorithm for the same reasons as in subject 2.

This subject indicated that he almost fell asleep twice. In the graph with the adjusted window size (blue) we see two peaks in Movement Density, which could indicate the movement that goes paired with arousal. For the red line (original threshold) this connection seems to be still there, but the peaks are not as high as with the blue line. When we look at the Heart Rate in each epoch, we do not see these peaks. The heart rate slowly decreases during the experiment, without peaks in Heart Rate when there are peaks in Movement Densities. This means that there is no connection between the Heart Rate and the result of the actigraphy algorithm.


Subject 3
(acc TH 0.15; HR mult. 1.96)
Movement Density
Heart Rate
Epoch Number


Subject 3
(acc TH 0.15; HR mult. 2.5)
Movement Density
Heart Rate
Epoch Number


Subject 3
(acc TH 0.1; HR mult. 1.96)
Movement Density
Heart Rate
Epoch Number


Subject 3
(acc TH 0.1; HR mult.2.5)
Movement Density
Heart Rate
Epoch Number