Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Distributed Computing

Smart Watch Actigraphy

Bachelor Thesis

Nicole Trachsler

tnicole@student.ethz.ch

Distributed Computing Group
Computer Engineering and Networks Laboratory
ETH Zürich

Supervisors:
Manuel Eichelberger, Philipp Brandes
Prof. Dr. Roger Wattenhofer

August 22, 2016

Acknowledgements

I thank Dr. sc. nat. Esther Werth from the Department of Neurology of the University Hospital Zurich. She was so kind to lent me two actimetry sensors and helped me interpret the gathered data from those two actimetry sensors. Furthermore, I thank my boyfriend Lukas for letting me track his sleep multiple times and for listening patiently to my concerns and giving supportive words.

Abstract

We use data of the accelerometer, gyroscope and heart rate sensor from a smart watch to classify the user's sleep into different sleep stages during the night. We process the gathered sensor data and implement a sleep stage classification algorithm.

All in this work compared sleep recording devices which are used by a sleep laboratory use only accelerometer data for the sleep classification, we however make also use of the data obtained from the heart rate sensor to achieve a more precise sleep stage classification.

We recorded 23 nights with two different subjects. Our sleep stage classification algorithm is implemented in Matlab and outputs several plots. We compare our results to the literature, two sensors from a sleep laboratory, to a fitness band from Polar and a fitness tracker from FitBit.

Contents

Chapter 1

Introduction

1.1 Motivation

In today's busy and labor-intensive world, many people have to deal with constant stress and often develop sleep disorders. Getting enough sleep of good quality during the night is important. The brain and the body need to regenerate itself during the night to start the day with a clear mind and full of energy. Sleep is key to a well-balanced body and mind.

To find out whether a person suffers from a severe sleep disorder, he has to visit a doctor several times and in the end he has to spend one or multiple nights in a sleep laboratory with polysomnography. Polysomnography is a diagnostic process to measure a person's sleep with multiple body parameters. A sleep scientist or doctor classifies every 30 seconds of the sleep into different sleep phases. This medical journey costs a lot of money and time, but it is the best available option to diagnose a sleep disorder as accurately as possible. If a person is not sleeping well lately - or is in general interested into his sleeping behavior - then actigraphy is the way to go. Actigraphy is a process used in sleep medicine and research to study a person's sleep-wake cycle. Data reflecting the person's activity level during the day and night are collected over a longer term with a so-called actimetry sensor.

Actigraphy has many advantages over polysomnography, it is cheap, no specialists are required and it can be done in one's own bed. The person has to wear only a small bracelet and there are not dozens of cables connected to his body like in a sleep laboratory during polysomnography. The person can stay in his familiar environment and his sleep is not changing because of an unfamiliar ambiance. Sleep recordings can be made over several days or weeks with no additional costs which gives access to a long-term overview of the person's sleep.

1.2 Limits

We only use accelerometer, gyroscope, and heart rate data. A sleep laboratory has real time access to many more sources of information about the person. The additional information of the sleep laboratory makes it possible to classify the sleep into different stages of sleep. Every person passes through multiple cycles of different sleep stages each night. Each sleep stage is clearly defined and has its own properties (e.g. sleep depth, eye movement). Thus our classification into the different sleep stages is limited from the beginning of our work, because of the limited information we get from our hardware. Therefore, actigraphy with a smart watch cannot replace polysomnography in a sleep laboratory.

Our smart watch actigraphy approach strongly depends on the availability of the built-in hardware. If it does not function, we cannot make any assumptions about the person's sleep during that period of time.

Furthermore, actigraphy with an accelerometer and gyroscope assumes that a lot of movement means the person is not asleep and low to no activity means that the person is asleep. This may hold for the majority of persons and recorded nights, but it is not true in all cases. Actigraphy therefore tends to overestimate the sleeping time. If a person lays restful in bed but is still awake, actigraphy may classify the person as asleep.

1.3 Goals

We aim to achieve a more accurate sleep classification than today's fitness trackers do. Most importantly, the classification will follow clear rules. People will be able to track their own sleep and have a first impression of whether their sleep behavior is normal or if they should consult a doctor. This avoids some unnecessary doctor's consultations. If a person thinks that something with his sleep is not in order, he can still consult a doctor. A doctor has then a closer look at the person's sleep and also takes into account possible influences of other disease.

We expect that most people who own a smart phone today, will wear a smart watch in the future. With our approach, these people will not have to wear any additional fitness tracker for sleep tracking anymore. It can all be done with a single smart watch. We also want to evaluate, whether smart watches can be used in medical actimetry in the future.

We want to make extensive use of the heart rate sensor to help distinguishing between the different sleep stages as good as possible.

Chapter 2

Sleep Basics and Related Work

This chapter provides sleep basics which are relevant for this work. It explains the different sleep stages and what sleep cycles are. The second part of this chapter discusses existing sleep detection devices and algorithms.

2.1 Sleep Stages

Rechtschaffen and Kales, two sleep researchers, declared six different stages of sleep (see Table 2.1) which are widely accepted and used in sleep medicine. REM stands for rapid-eye-movement and in this sleep stage the majority of dreams occur [1].

The sleep stages are categorized with the help of an electroencephalography (EEG), a polygraph and an electrooculograph (EOG). The EEG measures the electrical activity of the brain in voltage with electrodes attached to the person's head. It can differentiate brain waves of different frequency domains. A polygraph is a device which collects and displays data of various parameters of the person, for instance the blood pressure and the chest movement. The EOG measures the eye movements with multiple electrodes places around the person's eyes. The results of the EOG are mainly used to classify the REM phases.

2.2 Sleep Cycles

This section is mainly based on the chapter *Normal Human Sleep: An Overview* by Carskadon and Dement [1]. A healthy adult person passes through four or five sleep cycles every night. When a person falls asleep, he normally passes consecutively through stages 1, 2, 3 and 4. After he stayed in the deep sleep stage 4 for a while, he goes up again to stage 3, 2 and 1 until he reaches REM sleep. The person may wake up for a really short period before the next sleep cycle starts.

Table 2.1: Sleep stages classification according to Rechtschaffen and Kales [2].

Sleep Stage	Criteria	Physical State [3]
Awake	EEG contains alpha activity and/or low voltage, mixed frequency activity	Awake
REM	Low voltage, mixed frequency EEG together with episodic rapid-eye-movements	Dreaming
Stage 1	Low voltage, mixed frequency EEG without rapid-eye-movements	Very light sleep
Stage 2	12-14 Hz sleep spindles on the EEG, low voltage, mixed frequency EEG activity	Light sleep
Stage 3	Moderate amount of high amplitude, slow wave activity	Moderate deep sleep
Stage 4	Large amount of high amplitude, slow wave activity	Very deep sleep

Stage 4 is longer in the beginning of the night and usually is missing in the last cycle. On the other hand, REM sleep is short in the beginning and gets longer towards the end of the night. A visualization of the sleep cycles during a night is called a hypnogram. Figure 2.1 shows a typical hypnogram of the sleep of an adult person. Each cycle lasts around 90 minutes. In cycle 1 and cycle 2, the deep sleep phases are dominating the cycles, whereas in cycle 5 the deep sleep stage is not reached anymore, but the person stays in REM stage for almost one hour.

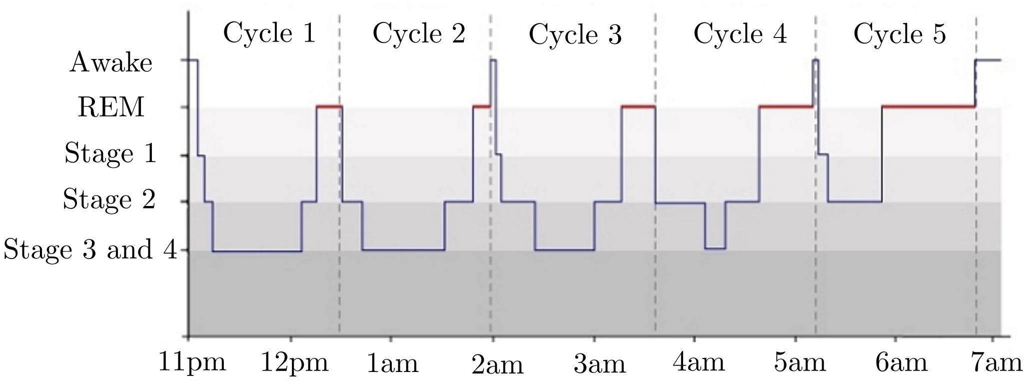


Figure 2.1: Oscillation between the different sleep stages during one night. Image by Luke Mastin [4].

The percentage a healthy adult person stays in one sleep stage is not equally distributed over the different sleep stages. In Table 2.2 is mentioned how much of the total sleep time each sleep stage makes up in percentage. These percentages may vary largely and are strongly depending on the person's age.

Table 2.2: Percentages per sleep stage.

Sleep Stage	Percentage
REM	20% - 25%
Stage 1	5%
Stage 2	45% - 55%
Stage 3 and 4	15% - 25%

2.3 Related Work

Nowadays there exist different fitness tracking devices which support sleep tracking from various commercial companies. Some of the popular companies are FitBit, Jawbone, Garmin and Huawei.

There already exist similar approaches like ours, but using the accelerometer of the smart phone to predict the user's sleep stage instead of the accelerometer of the smart watch. Some popular Android applications are *Sleep as Android* [5] and *Sleep Cycle alarm clock* [6]. Usually these applications include a smart alarm clock which claims to wake up the user when he is in a light sleep stage.

For these applications the smart phone must be placed on the user's bed. This has many disadvantages over sleep tracking with a smart watch. One disadvantage is that the user does not place the smart phone at the same position in his bed every night, so the sleep analysis may be different for the same night depending on the placement of the smart phone and thus intra-device reliability cannot be achieved. Also some people might not like it to have just a pillow between them and their smart phone. For our approach the smart phone can be placed further away from the bed within reaching distance of Bluetooth Low Energy.

A paper by Ancoli-Israel, Cole et al. evaluated the role of actigraphy in sleep research in 2003 [7]. They conclude that actigraphy is reliable in differentiating between sleep and wake periods for normal healthy adults. As soon as an adult suffers from a sleep disorder or has a disturbed sleep by any circumstances, the identification of sleep periods gets less reliable. The paper also points out that actigraphy is great in estimating the total sleep time but looses accuracy in identifying fragmented sleep.

A study by Chuan, Sheng and Xiaokang collected data of sleeping subjects simultaneously with a tri-axial accelerometer and a uni-axial accelerometer and the subject was also connected to a sleep monitor used as a reference standard [8].

They calculated the vector magnitude of every data point recorded by the tri-axial accelerometer. Then they feed the data sets from both accelerometers into a sleep scoring algorithm which estimates the sleep and wake periods. Finally, they compared the results to the sleep stages classifications of the sleep monitor. They detected that using a tri-axial accelerometer improves the sleep classification result compared to using a uni-axial accelerometer. Our approach also uses a tri-axial accelerometer in the smart watch and computes the vector magnitude of every data point.

In 1968 Baust and Bohnert found out that the changes of the heart rate associate to changes of the sleep stages [9]. We use this fact to improve our sleep classification algorithm.

Chapter 3

Sleep Classification Algorithm

This chapter describes our sleep classification algorithm in detail. The algorithm consists of three main parts: a movement classification, a heart rate classification and a sleep stage classification. Each part is described in a separate section.

3.1 Setup

We gather data with an LG G Watch R, using an application developed with Android Studio (version 1.5.1). The smart watch works with Android Wear and was released in 2014. The recorded data from each night is sent to a Motorola Moto G smart phone and saved into its external storage. Afterwards, we connect the smart phone to a computer and use Android Debug Bridge (ADB) to access the external storage of the smart phone and read out the saved data. On the computer the data is analyzed and evaluated with MATLAB (R2013a).

The application running on the smart watch gathers sensor data from the built-in accelerometer, gyroscope and heart rate (photoplethysmography) sensor. Data of these three sensors was gathered on the smart watch during 23 nights with a sampling rate of 0.2 seconds for every sensor. Every sample comes with an associated time stamp in milliseconds which is used to sort the samples chronologically and to perform analysis later.

3.2 Classification Intervals

To achieve a meaningful classification, we decided to classify short time intervals, for instance one minute, into the different sleep stages during the night. In Section 4.4 are the different lengths of the intervals evaluated. For each interval we look at the movement and heart rate classification and decide upon this information to which sleep stage it belongs.

3.3 Movement Classification

Data Preprocessing

The accelerometer sensor of the LG G Watch R has 3-axes and therefore we receive three values of the acceleration in x-, y-, and z-direction in $\frac{m}{s^2}$ for each sample. We calculate the difference between every two consecutive values of the x-, y-, respective z-direction and then take the norm.

$$\Delta_i^A = \left\| \begin{pmatrix} x_i - x_{i+1} \\ y_i - y_{i+1} \\ z_i - z_{i+1} \end{pmatrix} \right\| \tag{3.1}$$

This gives us $\overrightarrow{\Delta^A}$ for all accelerometer samples.

The gyroscope measures the rotation around the x-, y- and z-axis in $\frac{rad}{s}$ for each sample. With the gyroscope values we do the same calculations as in Equation (3.1) with the accelerometer values and end up with $\overrightarrow{\Delta^G}$.

Due to gyroscope data often not being available (see Section 4.1.2), we decided to use only the accelerometer data for the classification.

Unscaled Movement per Classification Interval

We iterate over all the classification intervals and every Δ_i^A within this interval is compared to a noise threshold. The noise threshold is determined with the smart watch recording sensor data for one hour while lying on a flat and stable surface. The recorded accelerometer data of one axis is expected to stay at the same value during the whole measurement. However, the data jumps, against our expectations, back and forth between two acceleration values with a distance of $0.1\frac{m}{s^2}$. Therefore, we set the noise threshold for our watch to $0.1\frac{m}{s^2}$.

For each interval we count how many Δ_i^A are bigger than the noise threshold and save the number into $moveCounter_i$.

Movement Categories

For the sleep stages classification described in Table 2.1, we are interested in distinguishing three movement categories: *high activity*, *low activity* and *no activity*. The classification into these three categories happens again per classification interval. We determine the limits between the three categories for every night recording dynamically, because some subjects tend to move more during sleep than others. We would get less accurate results if we would take the same limits for every recorded night and subject.

For the determination of the limits, we remove all $moveCounter_i$ which are zero. This is done, because we frequently have no data available (see Section 4.1.2) and therefore we cannot retrieve any information of the zero values. We build a box plot with all non-zero $moveCounter_i$ (see Figure 3.1) and retrieve the *median* and *upperWhisker* from the plot. The *upperWhisker* is by definition maximal 1.5 times longer than the box is long.

Table 3.1 states how we dynamically define the three activity categories. Each classification interval, for which data is available, is now categorized into one of the three activity categories.

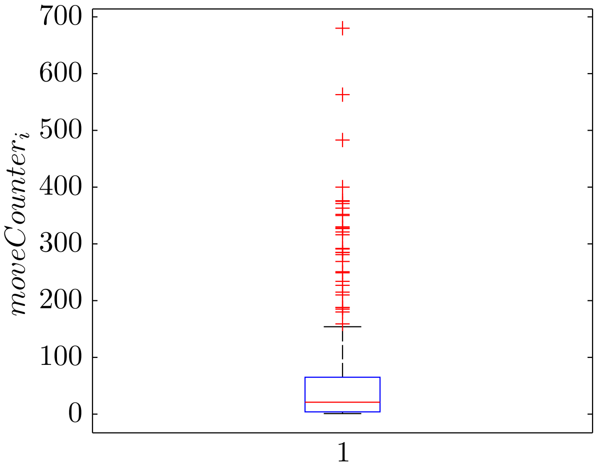


Figure 3.1: Box plot generated from all non-zero $moveCounter_i$ values from a sleep recording. The *median* of this box plot is at 21 movements per classification interval and the *upper whisker* at 154 movements per classification interval.

Table 3.1: Conditions on limits for the movement classification into the three different activity categories.

Activity	Condition
High activity	$moveCounter_i > upperWhisker$
Low activity	$upperWhisker \geq moveCounter_i > median$
No activity	$moveCounter_i < median$

3.4 Heart Rate Classification

The photoplethysmography sensor which measures the heart rate returns samples with an associated accuracy level between 0 and 3. Only heart rate measurements of accuracy level 3 are considered in this thesis, because the other accuracy levels are not reliable enough (see Section 4.1.1).

All heart rate samples with accuracy level smaller than 3 are filtered out and discarded. Since the heart rate samples tend to have large outliers, further description in Section 4.2.1, a smoothing algorithm is applied to get rid of these outliers. The smoothing algorithm shifts a sliding window with a specified span over the chronologically ordered samples and computes for each sample the average of all the heart rate samples which lie within the span. We use a time span of 121 seconds and evaluate in Section 4.2.1 our choice of this span. Figure 3.2 shows an example of a smoothed heart rate during a night recording.

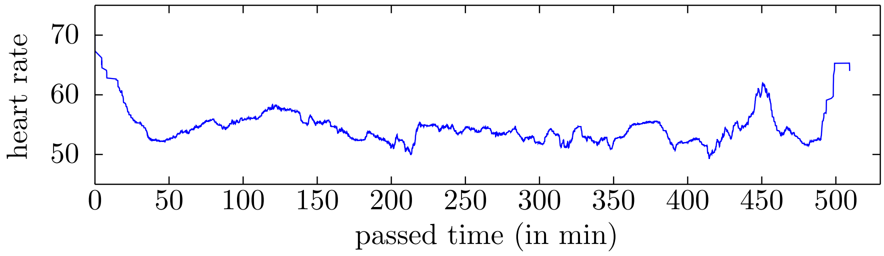


Figure 3.2: Example of a smoothed heart rate over one night.

After the smoothing is done, we iterate over all the intervals and look which heart rate samples belong to the current interval. Of all the heart rate samples belonging to one interval, the average heart rate is computed and rounded to an integer number.

3.5 Sleep Stages Classification

The preprocessed data of the movement and heart rate classification is used for the final sleep stages classification per classification interval.

Sleep Stages Aggregation

For the categorization into the six different sleep stages, defined by Rechtschaffen and Kales (see Section 2.1), the body movement can be used [3] as shown in Table 3.2.

The heart rate increases during REM sleep and stays low during stage 4 [4]. Nevertheless, there are no specific heart rate ranges for the different sleep stages, because the heart rate distribution is different for every person and can even change from one night to the other with the same subject (see Section 4.2.4).

Table 3.2 shows the movement and heart rate information that belongs to each of the six sleep stages. Stage 1 and 2 have exactly the same definition in terms of heart rate and movement. Therefore, it is not possible to distinguish between those two stages. We would need additional information from EEG, EOG and polygraphy to be able to distinguish between stage 1 and 2. The description of stage 3 and 4 is very close in terms of the physical state described in Table 2.1 and also in terms of the movement and heart rate classification.

For our classification, we decided to consider REM sleep individually as *REM*, stage 3 and stage 4 together as *deep sleep.* and the stages 1 and stage 2 are summarized into *light sleep.*

Table 3.2: Sleep stages classification with movement activity and heart rate.

Sleep Stage	Movement	Heart Rate
Awake	High	High
REM	Low	Alternating, often increased
Stage 1	Low	Below resting heart rate
Stage 2	Low	Below resting heart rate
Stage 3	No	Below resting heart rate
Stage 4	No	Lowest

Classification

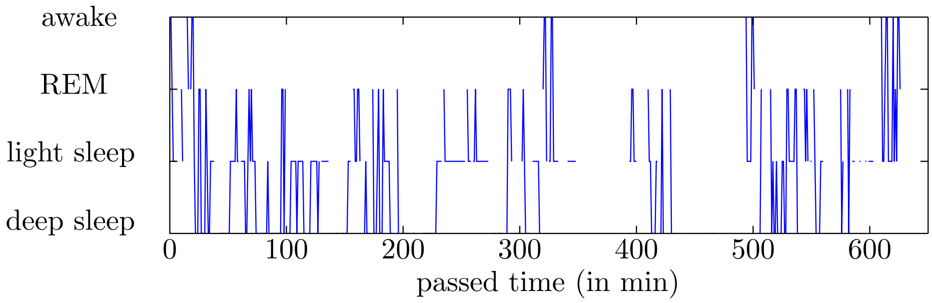
For the actual classification into the sleep stages, we iterate over the classification intervals one more time. The variable $meanHR$ stores the mean of all the heart rate values, where data is available and the variable $minHR$ stores the minimum value of those. The heart rate of the current interval i is accessed through HR_i. For each interval i, the algorithm performs the case distinction in Equation (3.2).

$$sleep\ stage_i = \begin{cases} deep\ sleep & \text{if } no\ activity \wedge HR_i < (meanHR - 1) \\ awake & \text{if } high\ activity \wedge HR_i > meanHR \\ REM & \text{if } low\ activity \wedge HR_i > minHR \\ light\ sleep & \text{otherwise} \end{cases} \tag{3.2}$$

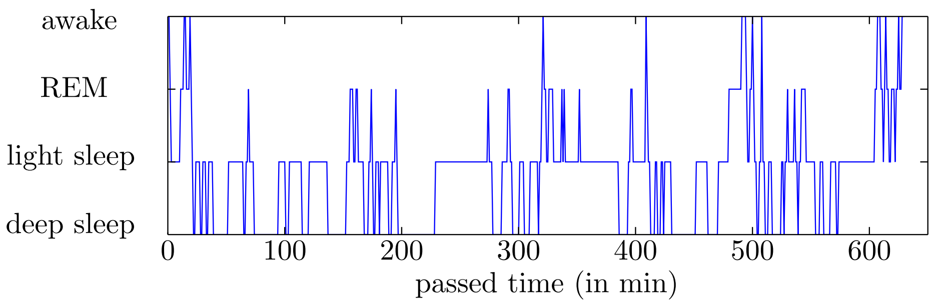
Each classification interval, for which data is available, is now categorized into one of the four sleep stages we consider in this work and specified in Section 3.5. A visualization is in Figure 3.3a.

Smoothing

To account for noisy input data with large outliers, we smooth the sleep stage classification. The smoothing function runs over the intervals of the sleep classification and computes for each interval the mean of itself and the two adjacent intervals on both sides, so it has a span of five intervals.



(a) Unsmoothed



(b) Smoothed with a span of five intervals.

Figure 3.3: Sleep classification of one night.

After the smoothing, the visual appearance of the sleep stages classification improves, as can be seen in Figure 3.3b. Small gaps for which no data is available due to erroneous sensors (see Section 4.1.2) are filled.

Chapter 4

Evaluation

This chapter discusses all the results and limitations of our actigraphy approach. Furthermore, it shows the evaluation of the sleep classification algorithm and compares our classification results to reports from devices of a sleep laboratory.

4.1 Samples

4.1.1 Data Accuracy

The heart rate, gyroscope and accelerometer sensors deliver with every sample an accuracy level between 3 and -1. Table 4.1 shows the meaning of the different accuracy levels [10].

Table 4.1: Accuracy levels of sensors in Android.

Accuracy Level	Meaning
3	Maximum accuracy
2	Average accuracy
1	Low accuracy
0	Unreliable
-1	No contact

The accuracy level of the gyroscope and accelerometer always stays at level 3, whereas the level of the heart rate sensor changes between level -1 and level 3 often during the night. When the subject is moving, especially in the beginning and the end of the measurement, the samples have mostly level 1 or 2. For a successful measurement of level 3 samples, the smart watch needs to be tightly fixated to the subject's wrist.

Figure 4.1 shows how less samples of level 3 are gathered after the subject woke up. The subject usually moves a lot more when awake than sleeping.

Due to the decreased accuracy (see Section 4.2.2) of the measurements of level 2 and 1, we only use heart rate samples of accuracy level 3 for this work.

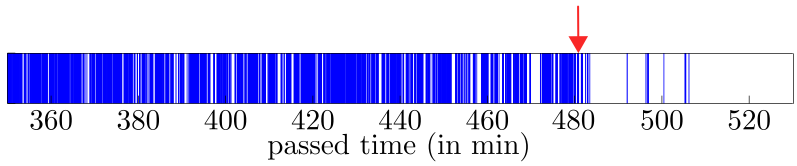


Figure 4.1: Availability of heart rate data with accuracy level 3. The red arrow indicates where the subject woke up and the blue bars show where heart rate data of accuracy level 3 is available.

4.1.2 Data Availability

The sampling rate of the gyroscope, accelerometer and heart rate sensor is set to 0.2 seconds and the sampling of all three sensors is turned on throughout the entire night. However, gaps because of missing data occur randomly in the recordings.

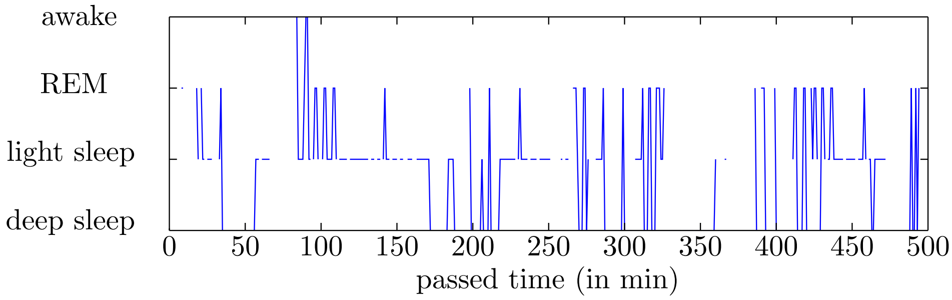
Looking at the data collected by the accelerometer and gyroscope sensor during all of our 23 recorded nights, accelerometer data is on average five times more frequent available than gyroscope data. Each movement which is detected by the gyroscope is also detected by the accelerometer. Therefore, the movement classification in this work is entirely based on the accelerometer data.

Table 4.2 shows how often which type of sensor data is available. The best availability of both relevant sensors during the night recording was 80% (see Figure 4.2a). Interrupts in the classification graph occur every time, when there is no data available for the current interval or only data from one of the two sensors. The worst case on the other hand is shown in Figure 4.2b with only 41% data availability of the heart rate and accelerometer sensor.

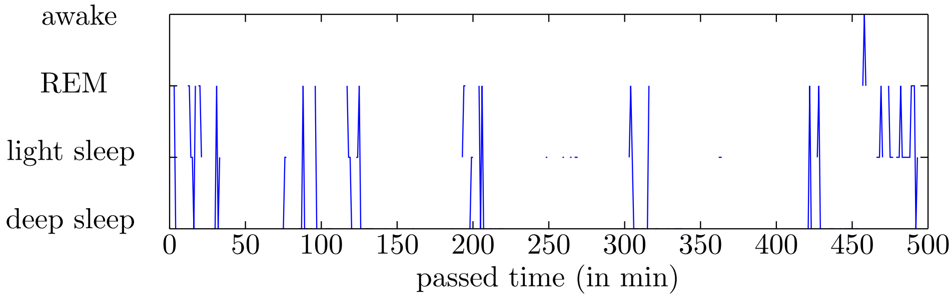
Table 4.2: Availability of sensor data per classification interval of one minute.

	Average Case	Worst Case	Best Case
Accelerometer	64%	56%	81%
Heart Rate	95%	66%	100%
Accelerometer and Heart Rate	62%	41%	80%

For a reasonable sleep classification, we need at least 65% of available data of both sensors. Missing data affects the sleep classification less, if it is uniformly distributed over the night recording. In contrast, when the missing data is clustered together, classification is not possible for this part of the sleep.



(a) Best case: Data availability of both sensors is 80%.



(b) Worst case: Data availability of both sensors is 41%.

Figure 4.2: Sleep classification of one night.

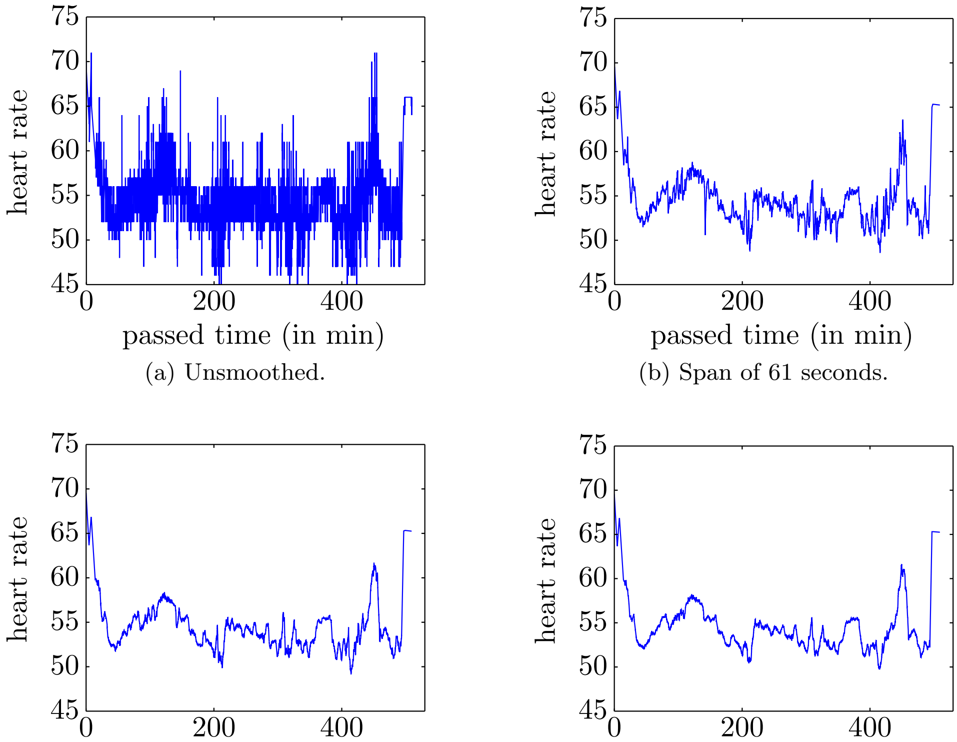
4.2 Heart Rate Data Comparison

We take the Polar H7 heart rate monitor as the ground truth for the heart rate measurements. The Polar H7 is a heart rate monitor in the form of a chest strap which transmits the data via Bluetooth Low Energy (BLE) to a paired device. The correlation of the Polar H7 to an electrocardiograph (ECG) is 99% [11]. ECGs are standard in medicine for recording the heart rate activity with electrodes attached to the subject's body. The electrodes measure the voltage generated by the contractions of the heart muscle.

4.2.1 Outliers and Noise

In Section 3.4 we used a span of 121 seconds for the smoothing of the heart rate during the night. Figure 4.3a shows raw heart rate data. There are multiple large outliers of the heart rate which are wrong due to erroneous measurement of the sensor. Furthermore, the signal also contains a lot of noise during the whole night. In Figure 4.3b, Figure 4.3c and Figure 4.3d different choices of the span are shown.

With a span of 121 seconds, the noise is eliminated and the oscillation of the heart rate over time is preserved. If the span gets bigger, more of the properties



(a) Unsmoothed. (b) Span of 61 seconds.

(c) Span of 121 seconds. (d) Span of 181 seconds.

Figure 4.3: Heart rate signal of accuracy level 3 during one night with different span lengths.

of the heart rate signal get lost.

Median filters are normally good in removing outliers from a signal, but in our case an average filter is the better choice. With a median filter, the signal gets corrupted at the beginning and in the end, where the heart rate signal is high because the subject is awake. Figure 4.4 shows that if a median filter with a span of 121 seconds is used, the heart rate is below 60 in the beginning of the measurement. The subject was still awake at the beginning and thus a heart rate below 60 is not possible. This happens, because we get significantly less accuracy level 3 samples when the subject is moving (see Section 4.1.1). Therefore, the median of the span lies more towards the middle of the night than towards the beginning respectively the end of the night.

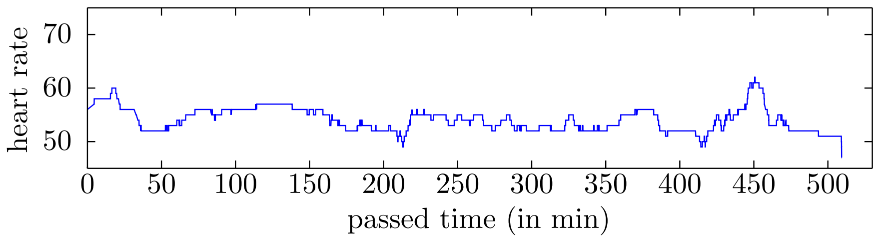


Figure 4.4: Heart rate signal of accuracy level 3 is processed with a median filter with span of length 121 seconds.

4.2.2 Heart Rate Sensor Accuracy

The root-mean-square error (RMSE) measures the difference of the heart rate values obtained from the smart watch compared to our ground truth, the Polar H7. Table 4.3 displays the RMSE values for the three accuracy levels. The first column shows the result obtained from calculating the RMSE with the raw heart rate data of both devices. Before calculating the RMSE in the second column, the values are smoothed with a time span of 121 seconds.

Table 4.3: RMSE of heart rate values between ground truth of Polar H7 and smart watch of one night.

Accuracy Level	RMSE	RMSE (smoothed)
3	2.78	0.82
2	5.36	2.66
1	8.35	4.19

The values for RMSE and RMSE (smoothed) are not exactly the same for each night, but the proportions are the same and RMSE (smoothed) with accuracy level 3 is always smaller than one. This concludes that the heart rate we are taking for the sleep classification is at most 1 heart beat off to the actual heart rate of the subject, because we also use 121 seconds smoothed data for the classification.

4.2.3 Cluster Anomaly

Figure 4.5 counts how often a specific heart rate occurs during one night. This histogram shows particularly, how clusters around specific heart rates are built for the LG watch. In each night measurement, these clusters form at the exact same heart rates (47, 52, 56, 61) and only the heights of the clusters differ. The Polar H7 on the other hand forms no clusters. Therefore, we assume that the cluster anomaly of the smart watch comes from either an erroneous heart rate sensor or the smart watch developers interpreted the digital signals of the heart rate sensor wrong.

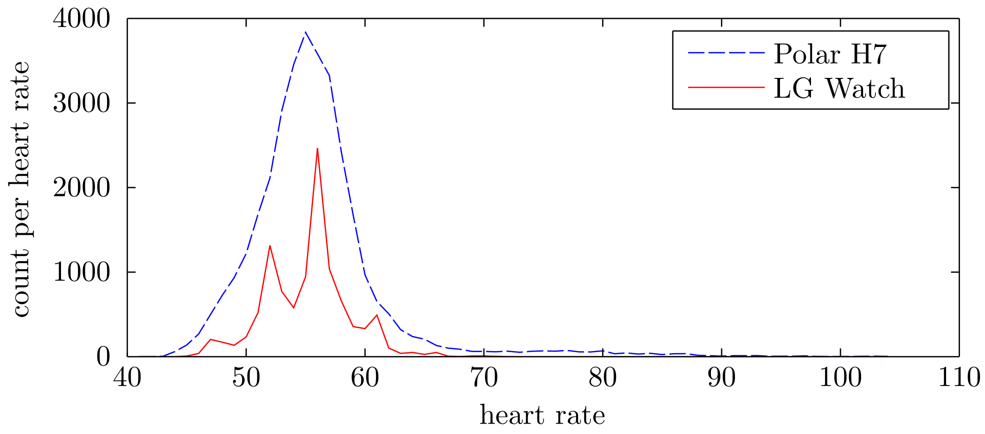


Figure 4.5: Histogram of the occurrences of the different heart rates of accuracy level 3.

4.2.4 Natural Heart Rate Variation

As mentioned in Section 3.5, the heart rate during sleep between different subjects is not equally distributed. Table 4.4 shows the mean heart rate during sleep of two subjects over three nights each. The first and last 30 minutes of the sleep measurement are not considered for the mean, because the subject may be awake at this time of the measurement and we are only interested in the heart rate values measured during sleep. Both subjects slept through these three nights and thus were not awake during the night.

Table 4.4: The values represent the mean heart rate of accuracy level 3 of two subjects during sleep.

	Subject A	Subject B
Night 1	49.2	61.2
Night 2	51.3	57.8
Night 3	49.6	55.3

Subject A has a mean heart rate around 50 beats per minute, whereas subject B's heart rate is up to 10 beats higher during sleep than A's heart rate. Also B's heart rate varies more between different nights. Therefore, it would not be a good design decision, to define static heart rate domains for the classification.

4.3 Sleep Classification Comparison

Polysomnography provides the only reliable ground truth for sleep classification, but it is very expensive. Due to this fact, we compare the results of our sleep classification algorithm to the literature for this thesis. We also have a look whether our result and the results of two actimetry sensors used in sleep laboratories are consistent. Furthermore, we discuss the sleep classification of the fitness tracker FitBit and compare it to our classification.

4.3.1 Literature

Sleep Stages

As already stated in Table 2.2, a healthy adult person stays in sleep stage 1 for 5%, in stage 2 for 45% to 55%, in stage 3 and 4 for 15% to 25% and in REM for 20% to 25% of the time. Transferred to our sleep stages definition, we end up with the theoretical percentages shown in Table 4.5. In the third column are the mean percentages we measured over all recorded nights and in the fourth column are the respective variances.

Table 4.5: Percentages per sleep stage used in this work and results we obtained.

Sleep Stage	Percentages in Theory	Measured Percentages	Variance
REM	20% - 25%	13%	0.18%
Light Sleep	50% - 60%	59%	0.52%
Deep Sleep	15% - 25%	28%	0.54%

The percentage measured for light sleep (59%) lies perfectly between the boundaries defined by the literature. On the other hand, our classification overestimates the deep sleep (28%) by a few percentages and underestimates the REM sleep (13%) clearly. The reason for the underestimation of the REM sleep may be that REM sleep has very different characteristics. Usually, the sleeping subject moves a little bit and the heart rate alternates (see Section 3.5). It is difficult to define what *low activity* means and thus explains our underestimation of this sleep stage. The most meaningful classification for REM sleep is however, as the name already states, the rapid-eye-movement which we cannot measure in our work because we only use heart rate and accelerometer data.

Sleep Cycles

As already mentioned in Section 2.2, a healthy adult subject has four or five sleep cycles during the night and each cycle last on average 90 minutes. In our sleep classification these cycles were always visible in nights with data availability over

60%. If a subject reported to have not slept well, there was usually significantly less deep sleep and the first sleep cycle started with a large delay.

In general, our algorithm assumes often that the subject is asleep in less than 10 minutes after the measurement has started, although this was only rarely the case in our 23 night recordings. Therefore the algorithm has difficulties to detect correctly when the subject falls asleep, because usually before falling asleep one shows the same behavior as in light sleep. Thus, the actual start of the first sleep cycle is not where our algorithm indicates it.

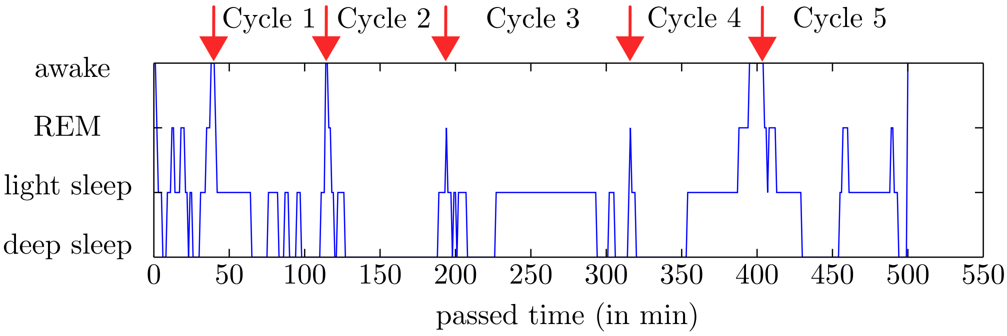


Figure 4.6: Sleep stage classification of one night. The beginnings of the sleep cycles are marked. Cycle 1 starts after 40 minutes and cycle 5 ends after 500 minutes.

Figure 4.6 shows a night recording with marked sleep cycles. The subject went through five sleep cycles during this night. The sleep cycles had a length of 80, 75, 120, 90 and 95 minutes which is on average 92 minutes. This meets almost exactly the expectation of on average 90 minutes per sleep cycle. The longest deep sleep phase lies in sleep cycle 2 and is thus in the first half of the night as expected.

4.3.2 Actimetry Sensors

In sleep medicine and research, actimetry sensors are widely used to get insight into the patient's movement habits. Patients wear an actimetry sensor over multiple days or weeks at home in their regular sleep environment. Usually the devices are waterproof, so the patient does not even have to take it off for showering.

A doctor or sleep scientist uploads the data from the device to a computer after the whole measurement period and runs the data through an evaluation algorithm from the company of the actimetry sensor. The obtained results reflect the patient' s movement behavior and the algorithm classifies the sleep and wake phases. Actimetry sensors cannot distinguish between different sleep stages and are therefore only used to study the patient's sleep-wake-cycle.

Actiwatch Spectrum

Actiwatch Spectrum is an actimetry sensor from the company Philips RESPIRONICS. It holds a 2-axis accelerometer, a light sensor and an additional sensor to detect whether or not the Actiwatch is worn on the wrist [12]. In the configuration phase before the measuring, an epoch length for the recording and later classification of the gathered data by the Actiwatch has to be chosen. An epoch is a short time interval of the whole recording over which the algorithm iterates and performs classification, equivalent to our classification intervals. We use an epoch length of one minute.

After the gathered data is uploaded onto a computer, a doctor or a sleep scientist has to identify *rest intervals* which show low activity manually. Within these *rest intervals*, the evaluation algorithm uses a *wake threshold* to define whether the subject is asleep or not. The algorithm sums up for every epoch within the *rest intervals* an *activity count* considering the neighboring epochs as defined in Figure 4.7.

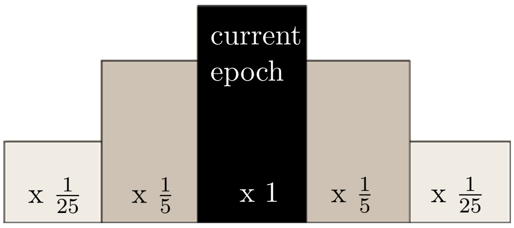


Figure 4.7: Definition of the calculation of the *activity count* for an epoch length of one minute [13].

If the activity count for an epoch is above the *wake threshold*, the subject is scored as awake, otherwise he is scored as asleep. The *wake threshold* can be set to different sensitivity levels by the sleep scientist, depending on the movement intensity of the subject during sleep [13].

Figure 4.8 shows the result of a night recording after the raw data run through the Actiwatch software and Figure 4.9 shows the movement measured and classified with the smart watch during the same night. There are several rest periods with almost no movement which are measured by both devices at the same time. Two examples of rest periods are between 32 - 92 minutes (22:11 - 23:11) and 195 - 252 minutes (00:54 - 01:51). The smart watch and the Actiwatch correlate also strongly with measuring high activity for example after 256 minutes (01:55), 320 minutes (02:59) and 500 minutes (5:49).

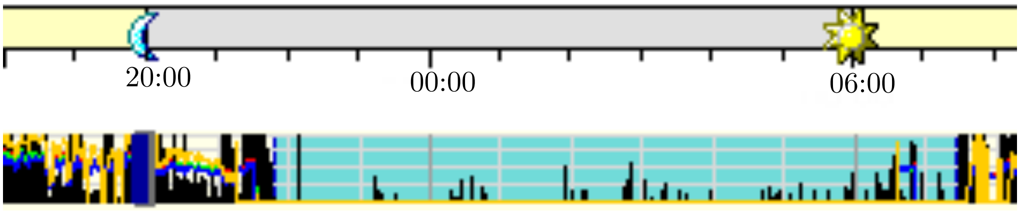


Figure 4.8: Night recording of the Actiwatch with corresponding time line. The light blue area indicates the resting period and the vertical black bars indicate the activity intensity. The yellow, red, green and blue curves are the light frequencies measured by the light sensor. The thick dark blue bar means that the Actiwatch was not worn on the wrist.

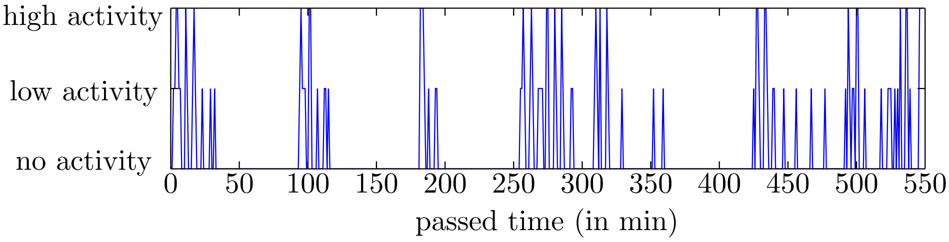


Figure 4.9: Movement measured and classified with the smart watch during the night.

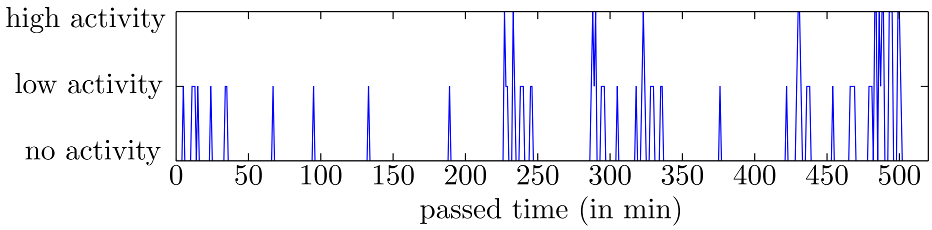
GENEActiv

GENEActiv is an actimetry sensor from the British company Activinsights. It includes a 3-axis accelerometer, a light sensor and a temperature sensor. It can measure with a frequency of up to 100Hz. For our work we set it to 10Hz which is the smallest available frequency [14].

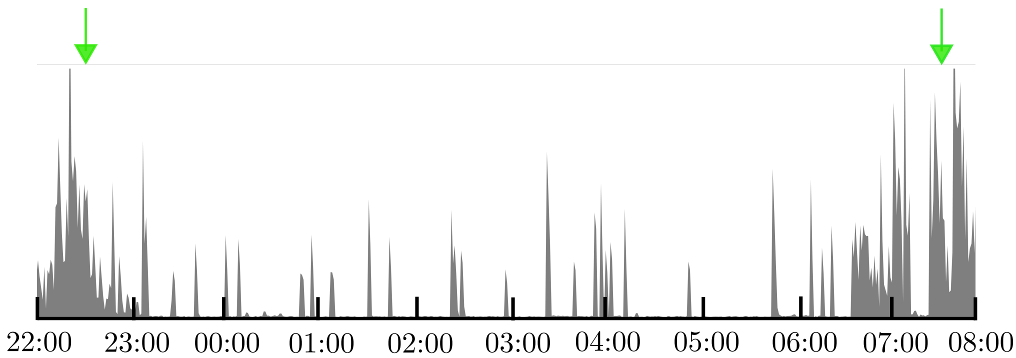
Activinsigts offer a free excel macro [15] to evaluate the gathered actimetry data from the GENEActiv. The macro produces activity charts, light & temperature charts and a summary of all recorded nights.

In Figure 4.10 we compare our movement data obtained with the smart watch to the activity chart from the GENEActiv. The measuring of the smart watch started at 22:34 and ended at 7:16. In Figure 4.10b the start and end time of the smart watch is indicated with green arrows. Both charts correlate strongly. When our algorithm detected high activity, there was also high activity in the chart of the GENEActiv for example after 227 minutes (02:21), 289 minutes (3:23) and 483 minutes (6:37).

The smart watch detected a rest period between 96 - 133 minutes (0:10 - 0:47) which is also visible in the GENEActiv activity chart. The GENEActiv often detects activity when we detected a rest period. For example between 134



(a) Movement measured by the smart watch during the night.



(b) Activity chart of one night recording by the GENEActiv produced by the free excel macro from Activinsights. The green arrows indicate the start respectively the end of the measuring period of the smart watch.

Figure 4.10: Movement classification of the smart watch and activity chart of the GENEActiv during the same night.

- 189 minutes (0:48 - 1:43) the smart watch detected the longest rest period of the night, but the evaluation of the GENEActiv shows four activity peaks during this period. The smart watch did not detect those activities because of unavailable data (see Section 4.1.2).

4.3.3 FitBit Charge HR

The FitBit Charge HR is a mainstream fitness tracker. It was explicitly designed for personal usage and not for medical purposes. In comparison to the actimetry sensors, the FitBit contains not only a 3-axis accelerometer, but also an optical heart rate sensor to continuously monitor the heart rate [16].

Automatic sleep tracking is only one of the things offered by the FitBit. As well as the actimetry sensors, the FitBit does not distinguish between the different sleep stages, but it distinguishes between restless and restful sleep and awake periods. When a subject was in a restful period with no movements and suddenly starts to move, the FitBit classifies this period as restless sleep. As soon as the movements get more intense, such that sleep is not possible anymore, the FitBit classifies the period as awake. Each minute is classified into one of these

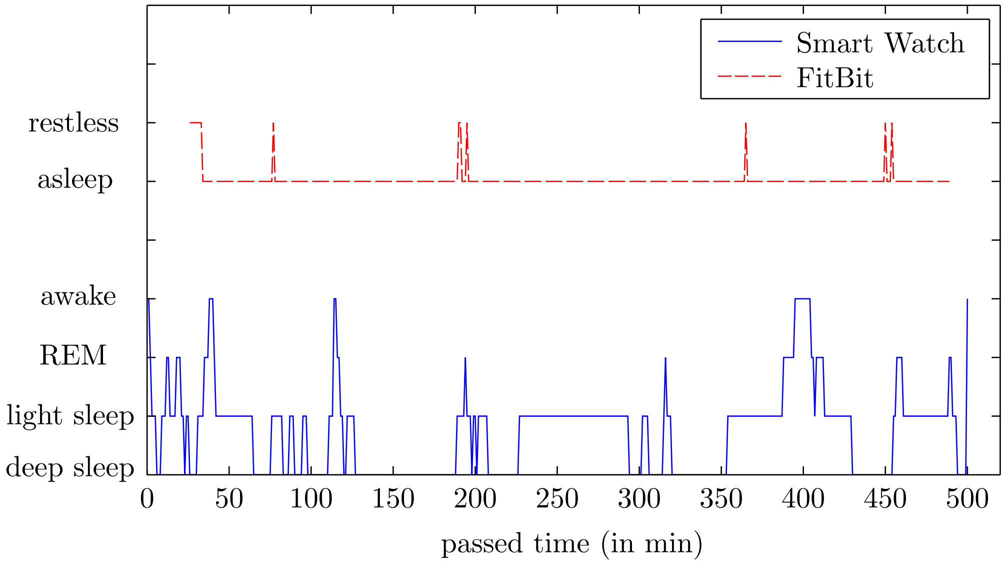


Figure 4.11: Classification of a whole night with the FitBit and the smart watch recorded simultaneously. The red dashed line shows the classification of the FitBit, when the line misses in the beginning and the end, the FitBit classified the subject as awake. The blue solid line shows the classification of the smart watch.

three types of sleep [17].

It is normal that a subject has short wake periods during the night as mentioned in Section 2.2. These wake periods usually last for under five minutes. Our tests show that the FitBit is not able to recognize short wake periods during the night. We measured five consecutive nights with the FitBit and the smart watch on the same wrist with the same subject and the FitBit did not report any short wake up period at all during the night. Figure 4.11 shows a night in which the subject reported to have woken up several times. The classification of the smart watch shows three short wake periods for this night after 40, 120 and 400 minutes. This is coherent with the subject's report. The FitBit on the other hand did not recognize any wake up of the subject during the night. The restless periods reported by the FitBit also did not occur at the same time as the wake periods by the smart watch were measured.

4.4 Classification Interval Length

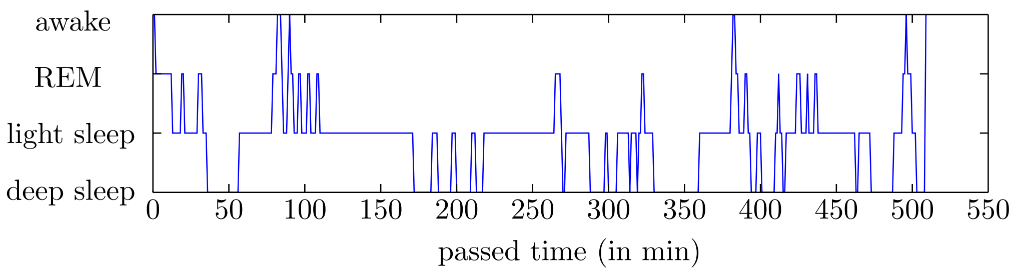
In Section 3.2 we explained the meaning of a classification interval. Figure 4.12 shows the sleep stages classification with our algorithm for three different interval lengths. The classifications are smoothed as described in Section 3.5 and the span used for smoothing decreases proportionally to the increasing classification interval length. With a classification interval length of one minute the span is five intervals, for a classification interval length of two minutes, the span is three intervals and for a classification interval length of five minutes no smoothing is done. Hence, in Figure 4.12c the data holes due to missing data are visible (at 380 and 480 minutes). In Figure 4.12a the sleep cycles are well visible, on the other hand in Figure 4.12c the visibility of the sleep cycles is not good anymore and therefore a lot of information is lost due to the bigger interval length.

We decided to do the sleep stage classification in our work for a classification interval length of one minute to not loose too much information. The majority of studies with actimetry sensors use classification intervals of length one minute [18].

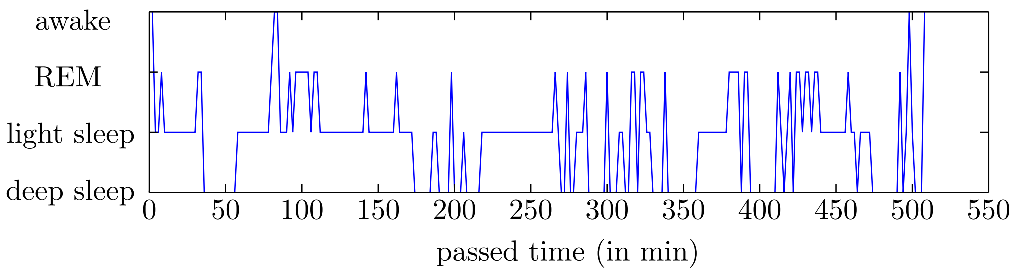
4.5 Battery Life

A recording of 8.5 hours during one night with the smart watch consumes about 35% of the battery. The screen is set to cinema-mode, to avoid it turning on automatically when the subject tilts his arm. Assuming the subject wants to use the smart watch for other purposes during the day, it must be charged once every day. A complete charging of the smart watch takes approximately one hour.

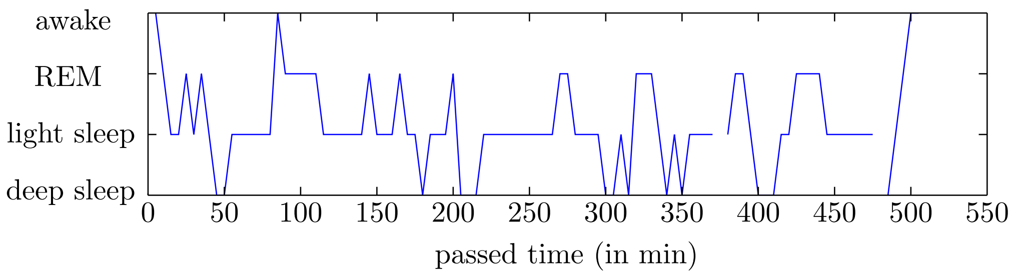
The battery life of the Actiwatch and GENEActiv depends on the adjustments made at the beginning of a new measurement. The Actiwatch lasts from 5 to 180 days, linear to the chosen epoch length. With an epoch length of one minute, it lasts for 22 days [13]. The GENEActiv lasts from 7 to 45 days, depending on the chosen measurement frequency between 100 Hz and 10 Hz [14].



(a) Classification interval length of one minute.



(b) Classification interval length of two minutes.



(c) Classification interval length of five minutes.

Figure 4.12: Sleep stage classification of one night with different classification interval lengths.

Chapter 5

Implementation

This chapter presents the implementation on the smart watch and smart phone. Due to time restrictions, there was not enough time to implement the sleep classification for the smart phone and design a nice graphical user interface. The sleep classification therefore runs entirely in Matlab.

5.1 Android Application

The Android application is implemented for Android Lollipop 5.0 (API level 21) which is supported by over 50% of the devices active in Google Play Store in July 2016 [19]. The application is called *Actigraphy* and consists of two parts: a smart phone application and a smart watch application. However, these two applications appear as one application which needs to be installed on the smart phone. The smart watch part is then transmitted from the smart phone to the smart watch automatically.

Figure 5.1 illustrates the states of the Actigraphy application. The arrows represent the data flow.

5.1.1 Smart Watch

The smart watch application consists of four classes. The two important classes are mentioned below. The third class is the MainActivtiy which is responsible for launching the application and the fourth class is a configuration class which keeps the constants organized.

The SensorService class runs always in the foreground. This ensures that the service is not killed by the operating system of the smart watch.

SensorService

The SensorService class has a method *onStartCommand* which can be activated by the user through a button press on the smart watch. This method registers listeners to the accelerometer, gyroscope and heart rate sensor. Another method *onSensorChanged* is called automatically every time a value of one of the three registered sensors changes. *onSensorChanged* reads out which sensor changed, what the new sensor values are and at what time the sensor values changed. The method then appends the read out values of one sensor change to a DataMap object. The DataMap is a data structure used to store information. It uses as a key the time at which the sensor change occurred and saves as a value the type of the sensor and the new sensor values. Before appending to the DataMap object, it is tested whether there is still space in the object. If it is full, the method *run* from the CommunicationWithHandheld class is called (see below).

The user can also press a button on the smart watch which activates the method *onDestroy*. This method unregisters the three listeners.

CommunicationWithHandheld

The CommunicationWithHandheld class contains a method called *run* which is called whenever the DataMap object is full. The method sends the full DataMap object over the data layer to the smart phone using Google Play Services. The Google Play Services manage the communication between the smart watch and the smart phone.

5.1.2 Smart Phone

The smart phone application has four classes MainActivity, ListenerService, SaveManager and Configuration. The MainActivity class launches the application and the Configuration class saves all constants. The tasks of the other two classes are mentioned below.

ListenerService

The ListenerService class contains a service which listens to incoming data from the smart watch. If a new DataMap object is received, the method *onDataChanged* is called automatically. *onDataChanged* iterates through every value in all the received DataMap objects and appends the sensor information, sensor values and time stamps to a StringBuilder. After the processing of one DataMap object, the method *saveFile* from the SaveManager class is called and the StringBuilder object is passed as an argument.

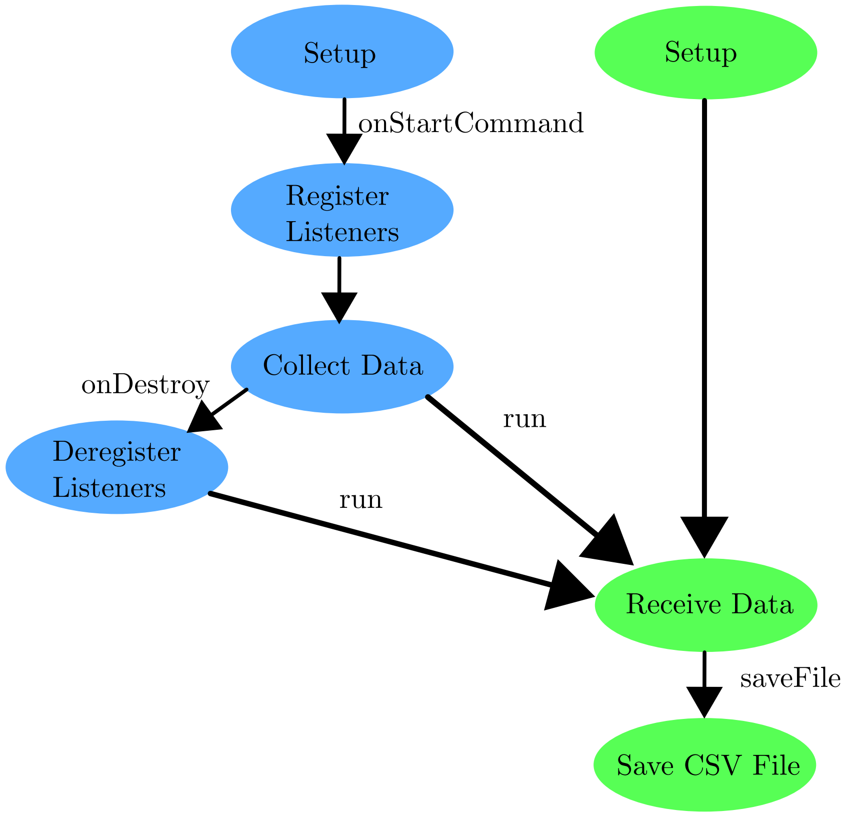


Figure 5.1: Data flow diagram of the Actigraphy application. The blue bubbles indicate states of the smart watch and the green bubbles indicate states of the smart phone. The text next to the arrows stands for the called methods.

This service is self managed by the operating system which means that the operating system automatically creates and destroys the service when needed.

SaveManager

The method *saveFile* from the SaveManager class takes the input StringBuilder object which contains the sensor data gathered by the smart watch and saves it in a comma-separated values (CSV) file into the smart phone's external storage. If *saveFile* was called before, there exists already a CSV file belonging to the current night recording. In this case the incoming StringBuilder object is appended to the already existing CSV file.

5.2 Matlab Scripts

The functions programmed in Matlab consist of three parts: The *movementDetector* implements the algorithm described in Section 3.3, the *heartRateDetector* follows the descriptions in Section 3.4 and the *sleepDetector* implements the sleep stages classification algorithm as described in Section 3.5.

A script with the name *detect* calls the three functions mentioned above one after the other with the correct input parameters and the CSV file of a chosen night recording.

Chapter 6
Conclusion

During our work, we discovered that the smart watch sensors suffer from cluster anomalies and data holes. These two factors weaken our sleep classification algorithm in such a way that it is often not possible to achieve a meaningful sleep stages classification for a night. Nevertheless, our movement results correlate strongly with the two actimetry sensors from the sleep laboratory.

The popular fitness tracker FitBit delivers a less accurate result of the sleep than our algorithm does. The FitBit is better in recognizing when the user falls asleep in the beginning of the night. Our algorithm is as good as the FitBit's algorithm in recognizing the wake up in the morning. During the night, we get much more accurate results with our sleep stage algorithm than with the FitBit. The FitBit is not able to distinguish between different sleep stages and as tests showed, it is also not good in recognizing wake up periods during the night.

Actimetry sensors' batteries hold long enough to gather information about the patient over more than a week, whereas smart watches need to be charged every day by the patient. Older patients may be over challenged with the charging of the smart watch. During the charging phase, no data can be collected and there is a chance that the patient forgets to put on the smart watch as soon as it is charged.

If in the future sleep laboratories use smart watches for actimetry, the already gathered data must be hidden from the patient during the measurement period. Otherwise if the patient can access his data immediately, he may change his behavior during the measurement period and for instance go to bed earlier than he used to go. This leads to biased data and therefore disturbs the diagnosing process of the sleep researchers and doctors.

Actimetry with a smart watch cannot replace polysomnography in any way. It provides a good informative basis about one's own sleep. People can use it to record their own sleep over long terms and it tells them whether they sleep enough or not. If someone wakes up not well recovered, he can check in his recordings whether he gets enough deep sleep for instance.

As soon as the hardware is ready, smart watch actigraphy is a great alternative to regular actimetry sensors. Tracking the heart rate along with the movement activity adds a new and valuable source of information to actigraphy for sleep medicine and research without requiring a specialized device.

6.1 Future Work

The sensors of the smart watch need to be improved, such that we get a continuous data stream from the smart watch with no data holes in the recording. It should not be a lottery, whether we gather enough sensor data or not during a night recording. Also the hardware specialists of the smart watch should have a closer look at the implementation of the heart rate sensor and find the reasons for the cluster building around specific heart rates during each night.

Our sleep stage algorithm should be compared directly to the gold standard in sleep research: polysomnography. With such a comparison, we would get exact results on how accurate our algorithm is.

One could gather actimetry data on a large-scale in the future with the smart watches. This would lead to a huge source of information about people's sleep. Sleep scientists could do large studies and statistics with this data.

Our algorithm tends to overestimate the sleep at the beginning of a night. A main improvement focus of the algorithm should be targeted at recognizing the beginning of the sleep, and thereby the first sleep cycle, more accurately. Also it should be studied why our sleep stage algorithm remains too long in deep sleep and how this could be changed.

A totally different approach would be to implement the sleep stage classification algorithm with the help of machine learning. Therefore one would need to collect raw sensor data of sleep recordings from many different subjects. For each of these sleep recordings one has to have an exact report about when the subject was awake or asleep. A machine can then be trained with the raw data and the accompanying sleep reports.

Bibliography

[1] Carskadon, M.A., Dement, W.C.: Normal Human Sleep: An Overview. In: Principles and practice of sleep medicine. 5th edn. Elsevier Saunders (2011)

[2] Rechtschaffen, A., Kales, A.: A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects. National Institute of Health, Los Angeles, United States of America (1968)

[3] Kambayashi, Y., Hagiwara, H.: Estimating Sleep Cycle Using Body Movement Density. In: 5th International Conference on BioMedical Engineering and Informatics, Chongqing, China. (2012)

[4] Mastin, L.: Types and Stages of Sleep. http://www.howsleepworks.com/types_cycles.html (2013) Accessed: 2016-07-02.

[5] Urbandroid Team: Sleep as Android. http://sleep.urbandroid.org/ Accessed: 2016-08-15.

[6] Northcube AB: Sleep Cycle alarm clock. https://www.sleepcycle.com/ Accessed: 2016-08-15.

[7] Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W., Pollak, C.P.: The Role of Actigraphy in the Study of Sleep and Circadian Rhythms. In: American Academy of Sleep Medicine Review Paper. (2003)

[8] Chuan, J., Sheng, Z., Xiaokang, L.: An Effective Way to Improve Actigraphic Algorithm by Using Tri-axial Accelerometer in Sleep Detection. In: 17th International Conference on Computational Science and Engineering. (2014)

[9] Baust, W., Bohnert, B.: The Regulation of Heart Rate During Sleep. In: Experimental brain research. Volume 7. (1968)

[10] Google Android: Sensor Manager. https://developer.android.com/reference/android/hardware/SensorManager.html Accessed: 2016-07-07.

[11] Altini, M.: Heart Rate Variability for Training. http://www.marcoaltini.com/blog/heart-rate-variability (2013) Accessed: 2016-07-07.

[12] Philips Respironics: Actiwatch Spectrum. http://www.usa.philips.com/healthcare/product/HC1046964/actiwatch-spectrum-activity-monitor Accessed: 2016-07-18.

[13] Respironics Inc.: Actiwatch Communication and Sleep Analysis Software (2008)

[14] Activinsights Ltd.: GENEActiv Instructions (2012)

[15] Activinsights Limited: GENEActiv Excel Macros. http://open.geneactiv.org/geneactiv_macros.htm Accessed: 2016-07-18.

[16] FitBit Inc.: FitBit Charge HR. https://www.fitbit.com/de/chargehr Accessed: 2016-07-18.

[17] FitBit Inc.: How do I track my sleep? https://help.fitbit.com/articles/en_US/Help_article/Sleep-tracking-FAQs#whatdothe Accessed: 2016-07-18.

[18] Berger, A.M., Wielgus, K.K., Young-McCaughan, S., Fischer, P., Farr, L., Lee, K.A.: Methodological Challenges When Using Actigraphy in Research. In: PubMed Central. (2009)

[19] Google Android: Android Dashboards. https://developer.android.com/about/dashboards/index.html Accessed: 2016-08-16.