

Spectral Analysis of Accelerometric Data to Identify Human Movement Patterns

Bachelor thesis

in the context of the academic studies in Medical Physics
at the Martin-Luther-University of Halle-Wittenberg
(Saxony-Anhalt, Germany)

presented by:

Patrick Wohlfahrt

tutor:

PD. Dr. rer. nat. Jan W. Kantelhardt

filig date:

21.08.2012



Martin-Luther-University of Halle-Wittenberg
Department of Physics

Contents

1	Introduction	5
2	Basic Principles and Methods	7
2.1	Physiological Basics	7
2.1.1	Accelerometry	7
2.1.2	Sleep	8
2.2	Investigated Data - National Cohort	10
2.3	Methods of Time Series Analysis	12
2.3.1	Fourier Transform	12
2.3.2	Logarithmic Binning	13
2.3.3	Moving Average	13
2.4	Statistical Methods	15
2.4.1	Receiver Operating Characteristics	15
2.4.2	Pearson Correlation	16
3	Results and Discussion	17
3.1	Spectral Analysis of Accelerometric Data	17
3.1.1	Frequency Spectrum of Whole Night	17
3.1.2	Power Spectrum Characteristics	19
3.1.3	Comparison between Day and Night	20
3.2	Comparison of Accelerometric Devices	22
3.3	Development of a Sleep-Awake Differentiation Algorithm	28
3.3.1	Parameter Selection	28
3.3.2	Classification Function	29
3.3.3	Algorithm Quality	32
3.3.4	Algorithm Test	36
4	Summary and Future Prospect	39
5	List of References	41
6	Statement of Authorship	45
7	Acknowledgement	47

1 Introduction

A human spends one third of his life sleeping, but he cannot consciously remember this time except maybe for a few dreams. Despite that, a human can judge whether or not he was sleeping well based on his daily physical fitness and mental balance. He has, hence, the ability to roughly assess his sleep quality. This subjective feeling is negatively affected by sleep disorders, such as apnea, hypopnea, or periodic limb movement. For this reason an objective sleep characterisation is preferred. Today, polysomnography is used almost exclusively to characterise sleep and to detect possible sleep disorders.

The necessity to sleep is still not fully understood. It is thus an active topic of research. Today's sleep research is committed to improving existing methods and developing new sleep classification concepts. One furthermore concentrates on the development of modern diagnostic tools for a better detection of sleep disorders. Another purpose is to develop an understanding of the risks and conditions associated with sleep disorders, to explore methods for reducing risks, and to exploit significant scientific findings for prevention practices. This is particularly important, because one quarter of the German population complains already about sleep disorders.

The German National Cohort, a network of German research institutions, targets the development of a new sleep analysis based on acceleration recordings measured by high-resolution accelerometers. This new diagnostic tool should be applied in a long-term epidemiologic study of 200,000 people starting in 2013 and then it should be established as a general method for analysing the human sleep behaviour. One hopes that results from the accelerometric analysis method are comparable to those obtained by polysomnography, thereby allowing for a more cost-effective assessment of sleep and sleep disorders. Another advantage is the size and the design of an accelerometer (comparable with a wrist watch) as well as its relatively long battery service life. This allows for continuous 7-day recordings from humans during their typical daily routine as well as during their natural sleep at home. In addition, new possibilities open up to detect specific human movement patterns and to reconstruct a person's normal daily routine.

This research work was conducted as part of the German National Cohort pre-study I and focused on identifying human movement patterns and, in particular, the development of a sleep-aware differentiation algorithm based on accelerometric recordings measured at subjects' wrists and hips. The formerly established sleep-aware differentiation algorithms as developed by Cole and Kripke in 1992 [1] or by Sadeh and Sharkey in 1994 [2] use very simple methods for identifying the human sleep behaviour, including arithmetic averaging, weighted moving averaging, and the standard deviation computation. The obtained results were promising, but the correct detection of wake phases was less successful. For this reason, this research work was based on spectral analysis which provides additional, possibly exploitable, parameters for sleep classification based on accelerometry.

The second chapter gives an insight into the physiological basics of human sleep and describes the functionality of an accelerometer. Moreover, the concept of the German National Cohort and the available accelerometric recordings are specified. Thirdly, the applied methods of time series analysis, including Fourier transform, logarithmic binning, and moving averaging, as well as the used statistical methods, such as receiver operating characteristics and Pearson correlation, are explained.

In the third chapter the obtained results are presented and discussed. At first the spectral analysis of accelerometric recordings is examined by means of power spectra and spectrograms. The differences in the human behaviour between day and night are also determined. During this research work three different accelerometric devices were used. Therefore, a detailed comparison between these accelerometers is presented, which provides guidance for

choosing the best device in the long-term epidemiologic study of 200,000 people. In the last part of the third chapter the development and test of a new sleep-aware differentiation algorithm is described.

Finally, a short summary and a prospect to future projects is given.

2 Basic Principles and Methods

To fully understand a complex phenomenon like sleep it is important to collect elementary information both about its physiological effectors as well as about possible means by which to measure and analyse. Typically, to analyse existing data, statistical techniques such as Pearson correlation and receiver operating characteristic (ROC) and physical methods, including Fourier transform, logarithmic binning, and moving averaging, are used.

2.1 Physiological Basics

This paragraph contains a short description of sleep characteristics as well as an illustration of important facts of accelerometry.

2.1.1 Accelerometry

Accelerometry is a non-invasive method of recording human movement and activity for prolonged periods. For an accelerometric measurement a small device - an accelerometer - which looks like a wrist watch, must be worn by a subject. Accelerometric devices are typically worn on the hip or the wrist of the non-dominant arm to prevent excessive measurements caused by unconscious movements.

Accelerometers measure frequency and intensity of body movements in terms of acceleration along a sensitive axis. It does not matter whether accelerations are caused by gravitation or translational and rotational body movements, or alternatively by soft tissue movements or external vibrations. All accelerometers use a specific form of spring/mass system (see Figure 2.1). When a force acts on a spring/mass system, the mass inside the device is accelerated, and the spring is consequently stretched or compressed. Deformation of the spring can be used to calculate the effective acceleration. Typical transducers are piezoelectric crystals, piezoresistive sensors, servo force balance transducers, and variable capacitance accelerometers [3].

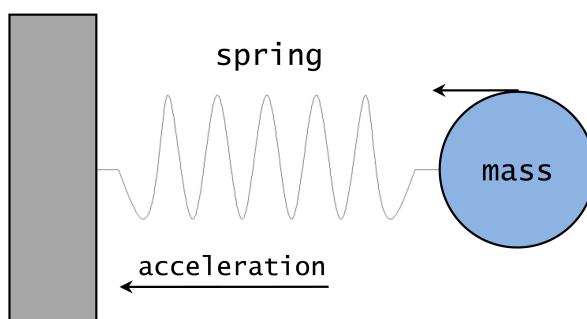


Fig. 2.1: Schematic diagram of a spring/mass system

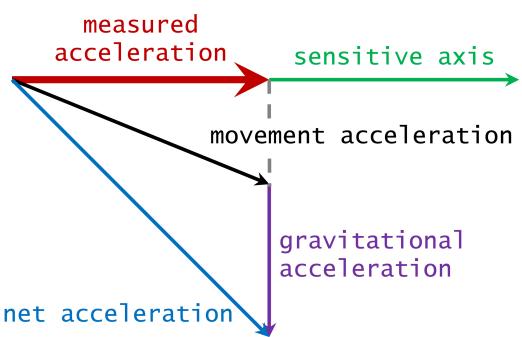


Fig. 2.2: Vectorial composition of measured signal

The structure of the measured acceleration of an accelerometer can be described by decomposition. When the device is motionless, only the gravitational acceleration vector is mapped onto the sensitive axis. When the device is moving, the measured acceleration is composed of the gravitational and movement acceleration as a vector sum in the direction of the sensitive axis (see Figure 2.2) [4].

If a device is worn too close to a rotation center or on a motionless area, the measured force might be weakened. Therefore choosing the right instrument/body position relative to the gravitational field and the intensity of movements helps to prevent or minimize possible signal error.

Accelerometry can, for example, be used to recognize typical patterns of movement during

sleep and sporting activities. To employ an accelerometer as a detector for specific movement features in an unsupervised way it is necessary to have an identification function or algorithm. In a number of medical and physical areas such an ability was found or will be explored in the next years.

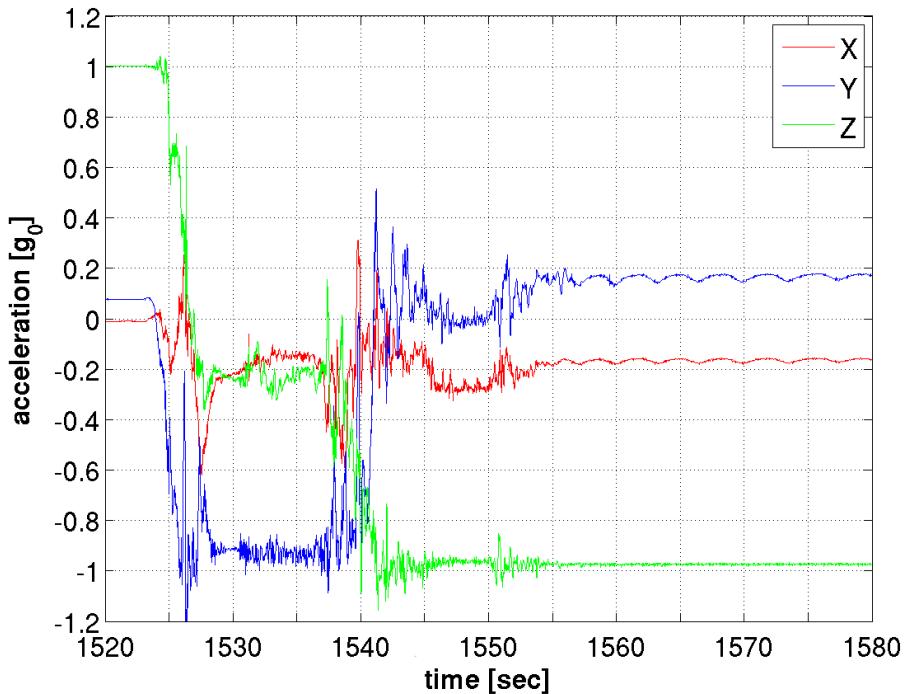


Fig. 2.3: Measured signal of a triaxial hip-mounted accelerometer during sleep in gravitational acceleration units g_0 (approximately $g_0 = 9.81 \frac{m}{s^2}$)

Figure 2.3 shows a typical measured acceleration signal of a triaxial accelerometric device. The periodic signal starting at a time of about 1557 seconds (especially in x and y direction) illustrates human respiration with a frequency of approximately 18 breaths per minute. In the time segment between 1525 and 1545 seconds strong acceleration fluctuations are shown, suggesting a stepwise turn over in bed by 180° (see reversed z direction of acceleration).

2.1.2 Sleep

Sleep is a naturally recurring state which differs from waking in many aspects, including a decreased or absent consciousness, comparatively reduced sensory activity and motor functions, as well as an inactivity of voluntary muscle movement [5]. A human spends one-third of his life sleeping. This time (on average, eight hours, daily) in a heightened anabolic state is needed to promote growth and to regenerate the immune, nervous, muscular and skeletal systems. Sleep is somewhat likened to coma or hibernation in that all are characterized by a reduction in reaction to stimuli; however, unlike coma or hibernation, reduced reactivity in sleep is far more easily reversed than are its coma or hibernation counterparts. The phenomenon of sleep is not well-understood by scientists, and is thus an active topic of research. A typical method to study sleep is polysomnography (PSG), wherein comprehensive recordings of biophysiological signals such as electroencephalography (EEG) of brain waves, electrooculography (EOG) of eye movements, electromyography (EMG) of skeletal muscle activity, electrocardiography (ECG) of heart beat and the respiratory flow for AHI detection, monitor many body functions, which are the basis of a sleep classification by polysomnography. Sleep is roughly divided into two categories according to physiological characteristics: rapid eye movement (REM) sleep and non-rapid eye movement (NREM) sleep. In 2007, the American Academy of Sleep Medicine (AASM) published a classification of NREM into three subgroups - stage N1 (light sleep), stage N2 (light sleep) and stage N3 (deep sleep) [6].

The main features distinguishing wakefulness, the three stages of NREM sleep, and REM sleep are shown in the following table.

Tab. 2.1: Main features of sleep and wake states (column 3 cited from [7])

stage	type	percentage	physiological	neurological
awake	wakefulness	< 5 %	normal body function, open eyes or closed eyes (relaxing)	β -waves [12 – 30 Hz] α -waves [8 – 13 Hz]
stage N1	light sleep	4 – 5 %	muscle activity ↓, twitching occasional muscles	ϑ -waves [4 – 7 Hz]
stage N2	light sleep	50 – 55 %	breathing pattern ↓, heart rate ↓, body temperature ↓	ϑ -waves [4 – 7 Hz], K-complexes, sleep spindles
stage N3	deep sleep	15 – 25 %	rhythmic breathing, limited muscle activity	δ -waves [1 – 4 Hz]
REM sleep	rapid eye mvt.	20 – 25 %	rapid eye movement, shallow breathing, relaxed muscles, heart rate ↑, blood pressure ↑	ϑ -waves, slow α -waves [≈ 8 Hz], β -waves

A night of a healthy human consists of four to five sleep cycles, each with a duration of 70 to 120 minutes. During a sleep cycle the different sleep stages are passed sequentially and the threshold to wake up increases. Additionally, the percentage of REM sleep in each consecutive sleep cycle increases, while deep sleep decreases. The stage REM is often known as dream sleep, because subjects report eventful dreams after waking from REM sleep. The final sleep scoring of the whole night is shown in a hypnogram (see Figure 2.4).

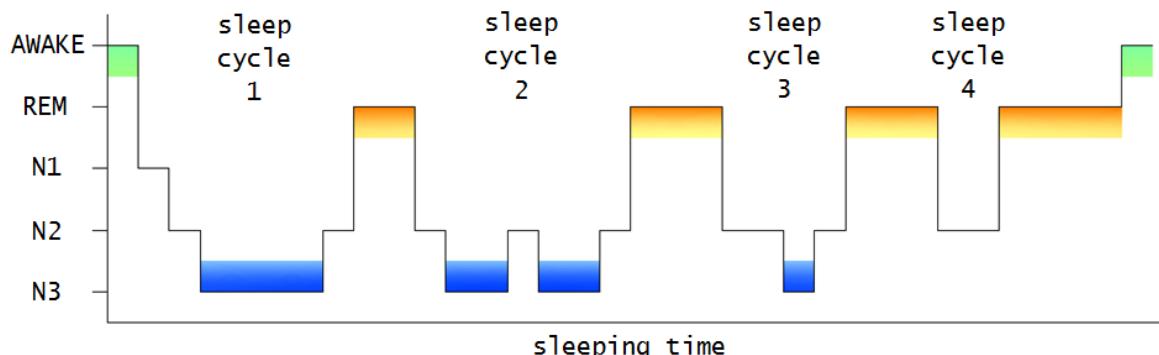


Fig. 2.4: Schematic hypnogram of a whole night of a healthy subject with four sleep cycles (deep sleep marked blue, REM sleep marked orange and wakefulness marked green)

On the basis of a polysomnography many sleep disorders, including apnea, hypopnea, and periodic limb movement (PLM) can be identified. An apnea is a respiratory arrest during sleep, which can be a cause of daytime sleepiness and microsleep. A reduced respiratory flow by at least 30 % over not less than 10 seconds is called hypopnea. Together, the count per hour, of those two symptoms of sleep apnea forms the Apnea-Hypopnea index (AHI). The AHI range is divided into 3 categories: 5 – 15 ≡ mild, 15 – 30 ≡ moderate and > 30 ≡ severe. Another sleep disorder, PLM, is characterized by spontaneous leg and arm movements. A typical parameter to gauge the severity of PLM is the number of limb movements per hour. All sleep disorders involve decreased sleep quality and a reduction in or absence of deep sleep, which is essential for physical relaxation.

2.2 Investigated Data - National Cohort

This research work was conducted as part of the German National Cohort pre-study I. The National Cohort consists of a network of German research institutions, including the publicly-funded Helmholtz and Leibniz Associations as well as numerous universities and research departments [8]. The objective of this research conglomerate is to develop a long-term epidemiologic study from mid-2013 to mid-2023 with particular emphasis on:

- Determining the cause of widespread diseases like cancer, diabetes and dementia,
- Understanding the relationships between lifestyle factors, environmental risk factors, chronic diseases and functional disorders,
- Gauging influences of geographical and socio-economic variations in health status and the risk of disease,
- Developing models for assessing the risk of chronic diseases and their prevention,
- Improving the early detection of diseases.

Within the scope of this study 200,000 people between the ages of 20 and 69 years (40,000 under 40 years) will be medically examined and interviewed about their personal lifestyle habits. The study participants will be selected to represent the percentage of all age and population strata in Germany. Furthermore, smaller subgroups of 40,000 representative people will be formed to make some additional investigations, such as a whole-body, heart and brain magnetic resonance imaging (MRI), a 24-h ECG, sleep apnea assessments as well as enhanced motoric, sensoric and cognitive function tests. Five years after the baseline assessment all participants will be reinvited for a second medical examination, which will include the baseline program.

Currently the final tasks of the preparatory phase no. 1 have been completed. Besides the planning of the epidemiologic study a feasibility study and a pilot study of 2,000 people have been conducted. These two preliminary tests are divided into many subgroups with different examination fields. The basis for this research project is to amass data sets in the following areas [9]:

- P4: Physical Activity-Group
 - Evaluation of the feasibility of mono- and multifunctional devices to objectively measure physical activity
 - Examination centers: Berlin, Regensburg, Kiel, Heidelberg, Bremen
 - Number of participants: 270
 - Used devices: Actigraph GT3Xplus, GENEactiv, SOMNOWatch
- P6: Sleeping-Group
 - Multipurpose sleep assessment module
 - Examination centers: Halle (Saale), Greifswald, Bremen, Münster
 - Number of participants: 100
 - Used devices: Actigraph GT3Xplus, SOMNOWatch (localised to wrist & hip)

The P4 study comprises activity measurements for a period of seven days and the P6 study includes a polysomnographic recording and activity measurements for one night as well as a three-day activity measurement with a SOMNOwatch device worn on the wrist. The one night measurements of the P6 study are used to develop a sleep-aware-differentiation (SAD) algorithm in comparison to PSG and the individual software tools of the device manufacturers. For evaluating the three main devices Actigraph GT3Xplus, GENEactiv, and SOMNOwatch all data sets are consulted.



Fig. 2.5: Three main accelerometers SOMNOwatch, Actigraph GT3Xplus, and GENEactiv (from left to right)

The main features of the three accelerometers used, Actigraph GT3Xplus, GENEactiv, and SOMNOwatch (see Figure 2.5), are summarised in Table 2.2.

Tab. 2.2: The main features of the accelerometers Actigraph GT3Xplus, GENEactiv, and SOMNOwatch (according to [9])

features	SOMNOwatch	Actigraph GT3Xplus	GENEactiv
triaxial	yes (range: $\pm 6g$)	yes (range: $\pm 6g$)	yes (range: $\pm 8g$)
raw data	yes	yes	yes
sampling rate	128 Hz	80 Hz	100 Hz
battery life	7 days	8 days	7,5 days
storage capacity	256 MB	256 MB	400 MB
size [mm]	45x45x16	45x34x19	43x40x13
weight	30 g	19 g	17 g
placement	hip & wrist	hip	hip
software (sw)	SOMNOwatch sw	ActiLife5	GeneActiv sw
software price	890€	333€	N/A
device price	1250€	240€	213€

2.3 Methods of Time Series Analysis

This section gives an overview of physical and mathematical practices that are used in this work.

2.3.1 Fourier Transform

An often used method in physics is the Fourier transform \mathcal{F} . This mathematical transform is mostly employed to convert a function of time $f(t)$ into its frequency spectrum (function of frequency $g(\omega)$). With the inverse Fourier transform \mathcal{F}^{-1} one gains the function of time from the frequency spectrum.

The mathematical definition of the Fourier transform of non-periodic and integrable functions $f : \mathbb{R} \rightarrow \mathbb{C}$ over an infinite interval is

$$\mathcal{F}[f(t)] = g(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt \quad (1)$$

$$\mathcal{F}^{-1}[g(\omega)] = f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(\omega) e^{i\omega t} d\omega. \quad (2)$$

To analyse the frequency spectrum of a measured time series with a length T the implementation of the discrete Fourier transform (DFT) is needed:

$$g(n\Delta\omega) = \sum_{m=0}^{N-1} f(m\Delta t) e^{-2\pi i nm/N} \quad (3)$$

$$g(k\Delta t) = \frac{1}{N} \sum_{n=0}^{N-1} g(n\Delta\omega) e^{2\pi i kn/N} \quad (4)$$

$$\text{including } \Delta\omega = 2\pi/T, \quad \Delta t = T/(N-1)$$

Each of the DFT equations is calculated N times, which indicates two nested loops with an individual size of N . Consequently a DFT of N measured values demand $\mathcal{O}(N^2)$ operations. Hence with an increasing N the DFT becomes very slow. A more efficient algorithm of DFT is the fast Fourier transform (FFT). An FFT can be applied if the counts of measured values N can be expressed as power of two. By means of an FFT the number of operations is reduced to $\mathcal{O}(N \log_2 N)$ and the results are calculated $\frac{N}{\log_2 N}$ times faster.

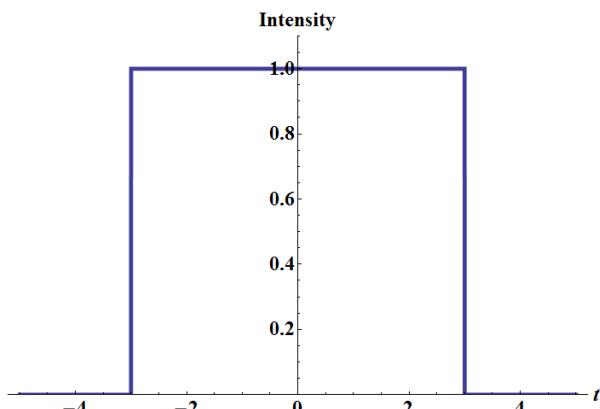


Fig. 2.6: Function of time $f(t) = \vartheta(t+3) - \vartheta(t-3)$

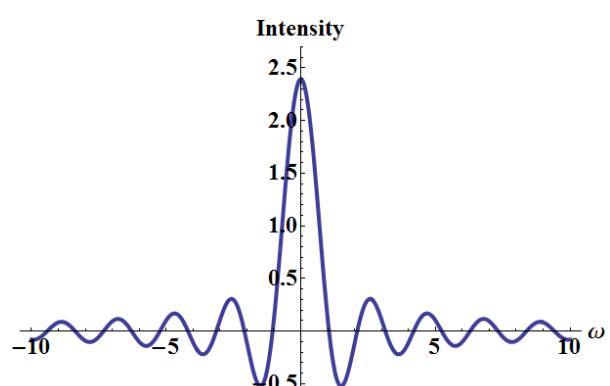


Fig. 2.7: Function of frequency $g(\omega) = \sqrt{\frac{2}{\pi}} \frac{\sin(3\omega)}{\omega}$

2.3.2 Logarithmic Binning

Logarithmic binning is a method to calculate the logarithmic average of a specific percentage p of N data points. In the context of this work logarithmic binning is used to improve the illustration of a double logarithmic power spectrum.

The number of data points after i iterations n_i can be calculated according to the formula

$$n_i = \lceil (1 - p) n_{i-1} \rceil, i \in \mathbb{N} \quad (5)$$

where n_{i-1} ... number of data points of bin start

n_i ... number of data points of bin end

$1 - p$... binfactor

With the aid of an FFT the logarithmically binned argument x_i and functional value $f(x_i)$ are determined as follows:

$$x_i = \log_{10} \left(\frac{n_i + (n_{i-1} + 1)}{2} \cdot \frac{1}{N} \cdot \frac{\nu}{2} \right) \quad (6)$$

$$f(x_i) = \frac{1}{n_{i-1} - n_i} \left(\sum_{k=n_i+1}^{n_{i-1}} \log_{10} |g(k\Delta\omega)|^2 \right) \quad (7)$$

where N ... total number of data points

ν ... sampling rate of accelerometric device

$g(k\Delta\omega)$... discrete frequency spectrum

Figure 2.8 and 2.9 reveal the differences between a binned and an unbinned power spectrum. It is apparent that a binned, double logarithmic power spectrum illustrates more representative information than an unbinned one.

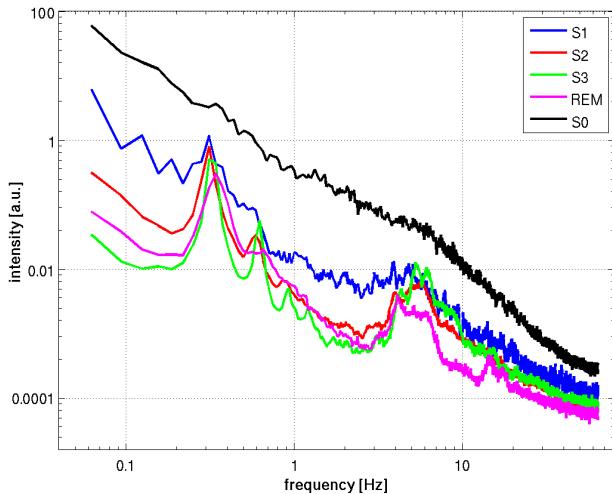


Fig. 2.8: Unbinned double logarithmic power spectrum

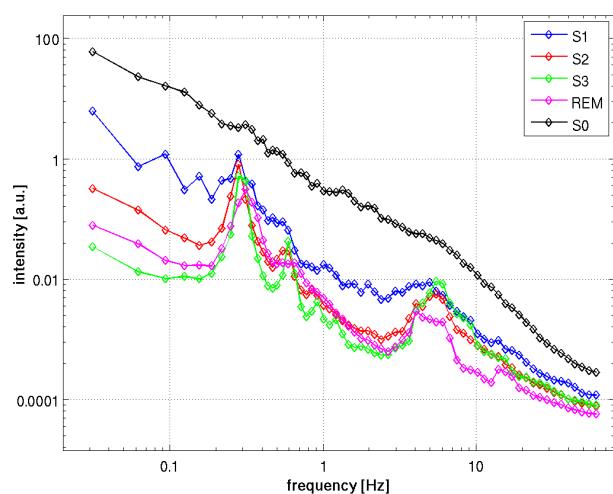


Fig. 2.9: Binned double logarithmic power spectrum

2.3.3 Moving Average

A moving average is a kind of smoothing to analyse time series by generating an array of averages of different parts of the entire signal, thereby emphasizing long-term trends and removing short-term fluctuations. The division into long- and short-term events depends on the selected averaging method and the chosen timeframe.

In general this method employs a repeating average calculation over fixed subsets until the end of the time series is reached. Specifically, the first moving average is determined by the values of the prime fixed subset. Following the initial averaging the fixed subset is moved

one value forward and the moving average of the new subset is calculated. This procedure continues until the subset includes the last time series value.

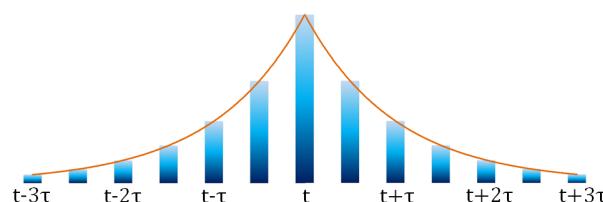
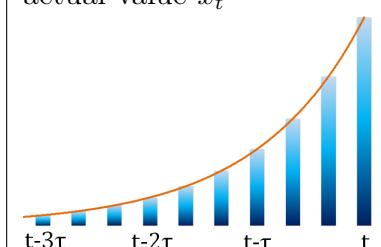
Let $\{x_t\}_{i=1,\dots,N}$ denote a time series and let τ be the oneside timeframe of a centered fixed subset, then the elements of the simple moving averaged time series are (almost always) given by

$$\tilde{x}_t = \frac{1}{2\tau + 1} \sum_{n=-\tau}^{\tau} x_{t+n} . \quad (8)$$

A centered simple moving average (SMA) after the formula (8) will only exist if $\tau + 1 \leq t \leq N - \tau$. One might either ignore the undefined elements or use weighted averages of corresponding smaller incomplete windows instead.

Table 2.3 shows two methods of exponential moving average (EMA) which were used to develop a sleep-aware-differentiation algorithm (SAD algorithm).

Tab. 2.3: Two different methods of exponential moving average (EMA)

features	peaked EMA	backward EMA
exp. precision	$\lambda = \begin{cases} t, & \text{if } \tau \leq t < 3\tau \\ N-t, & \text{if } \tau \leq N-t < 3\tau \\ 3\tau, & \text{if } t \geq 3\tau \wedge N-t \geq 3\tau \end{cases}$	$\lambda = \begin{cases} t, & \text{if } \tau \leq t < 3\tau \\ 3\tau, & \text{if } t \geq 3\tau \end{cases}$
formula	$\tilde{x}_t = \frac{1}{2\tau} \sum_{n=-\lambda}^{\lambda} x_{t+n} e^{- n /\tau}$	$\tilde{x}_t = \frac{1}{\tau} \sum_{n=0}^{\lambda} x_{t-n} e^{-n/\tau}$
illustration	actual value x_t 	actual value x_t 
essential data	x_t, λ former and λ future values	x_t and λ former values

2.4 Statistical Methods

With statistical principles it is possible to estimate the quality of a developed algorithm and physical correlations.

2.4.1 Receiver Operating Characteristics

Based on a binary classification like a sleep-aware differentiation (SAD) the performance of a medical test or algorithm can be evaluated by using sensitivity (Sn) and specificity (Sp). These two quantities are determined by comparing the test result and the real value. The test result depends on a discrimination threshold in order that a score below the threshold means sleep and a score above the threshold means awake. A variation of the discrimination threshold implies a change of the test result, which is combined with a modification of sensitivity and specificity. However, in general such a threshold will not perfectly distinguish two states. To find the optimal threshold a plot of sensitivity versus specificity at various threshold settings is used. This graphical illustration is called receiver operating characteristic (ROC) curve.

	PSG AWAKE	PSG SLEEP
MEASURE AWAKE	counts: ww SENSITIVITY (Sn)	counts: sw FALLOUT (Fo)
Positive Predictive Value (PPV)		
MEASURE SLEEP	counts: ws MISS RATE (Mr)	counts: ss SPECIFICITY (Sp)
Negative Predictive Value (NPV)		

Fig. 2.10: Diagnostic indices for SAD

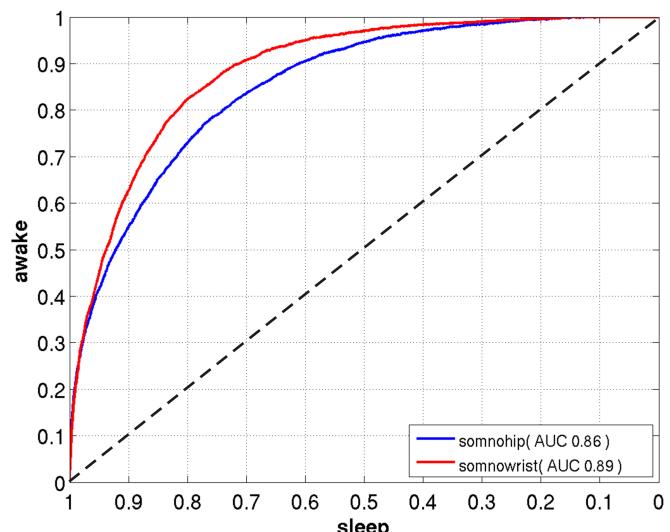


Fig. 2.11: ROC of a SAD with accelerometry

Taking polysomnography (PSG) based values to be the true values, a ROC-type analysis can be applied to sleep-aware differentiation. One compares the test result and the true values of a subject for obtaining the number of properly classified awake states (true positive value - ww) and the number of properly classified sleep states (true negative value - ss) as well as the number of falsely identified awake states (false positive value - ws) and the number of falsely identified sleep states (false negative value - sw). Evaluating the quality of the test result the sensitivity (Sn) and the specificity (Sp) are defined as probabilities of correct awake scoring and correct sleep scoring. Additional probabilities such as the positive predictive value (PPV) or the negative predictive value (NPV) describes the number of proper awake or sleep scores relative to the total number of awake or sleep scores. The relationships between these diagnostic indices are shown in Figure 2.10 and are described, mathematically, with the following formulas:

$$\begin{aligned} Sn &= \frac{ww}{ww + ws} & Sp &= \frac{ss}{ss + sw} & Mr &= 1 - Sn \\ PPV &= \frac{ww}{ww + sw} & NPV &= \frac{ss}{ss + ws} & Fo &= 1 - Sp \end{aligned}$$

A pair of Sn and Sp reflects diagnostics accuracy more usefully than does a single point, because an individual analysis enables the study of only one real value, either awake or sleep.

The combination of independent variables, however, enhances the statistically evaluable amount of data and doubles the range of possible results [10]. The performance of a developed method can be illustrated in a ROC curve with the correct sleep scoring on the x-axis and the correct awake scoring on the y-axis (see Figure 2.11). All values above the dividing diagonal display better classification results than random and the point (1,1) is the aim of a perfect classification.

One of the most established quality criterion of a diagnostic test is the area under the ROC curve (AUC). The AUC is calculated by using the upper sum U or lower sum L method. To determine precise results one must use a large quantity of discrimination thresholds.

$$U := \sum_{i=1}^n (x_i - x_{i-1}) \cdot \max \{y_{i-1}, y_i\} \quad L := \sum_{i=1}^n (x_i - x_{i-1}) \cdot \min \{y_{i-1}, y_i\} \quad (9)$$

The AUC can assume values ranging from 0 to 1. An area of 1 means a perfect classification, because all sleep and awake states are scored correctly for one discrimination threshold ($Sn = Sp = 1$). The worst case is an AUC of 0.5 where the test will not be better than random [11]. If the AUC is obvious less than 0.5, an inverted scoring will reach a good quality, i.e., either the test result has to be negated or the scoring above or below the threshold (awake or sleep) has to be swapped.

2.4.2 Pearson Correlation

The Pearson product-moment correlation coefficient r_{xy} is a measure of a linear dependence between two normally distributed quantities x and y and is normalised to values from -1 to 1 [12].

If x and y are independent random variables the coefficient $r_{xy} = 0$. A linear equation $y = mx + n$ best fits the correlation between x and y if all ordered pairs of data points (x, y) lie on a straight line. In this case, $r_{xy}^2 = 1$. A value of r_{xy} greater than zero implies an increase of x and a simultaneous rise of y , and a value less than zero is indicative of an increase of x and a concurrent decrease of y .

The mathematical definition of the Pearson correlation coefficient, r_{xy} , is the covariance of x and y divided by the product of their standard deviations, σ_x and σ_y .

$$r_{xy} = \frac{Cov(x, y)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^n (x_i - \langle x \rangle)(y_i - \langle y \rangle)}{\sqrt{\sum_{i=1}^n (x_i - \langle x \rangle)^2} \sqrt{\sum_{i=1}^n (y_i - \langle y \rangle)^2}} \quad (10)$$

To reduce the calculation time of r_{xy} the formula is transformed by noting that a previous averaging of x and y is not necessary to determine r_{xy} .

$$r_{xy} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \sqrt{n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}} \quad (11)$$

3 Results and Discussion

With the aid of the content-related basics and the methodological tools the accelerometric data can be examined. Thereby the spectral analysis serves as a basis of this research.

3.1 Spectral Analysis of Accelerometric Data

The spectral analysis is a popular method, which determines how strongly a frequency range is represented in a time. The discrete Fourier transform is used to calculate the frequency spectrum of a discrete time signal.

3.1.1 Frequency Spectrum of Whole Night

Human sleep behaviour is related with numerous physiological processes, including muscle activity, breathing, blood pressure, and heartbeat. A variation of some body functions is associated with a change of sleep stage [13]. In particular, the muscle activity reduces significantly in intensity and in frequency of occurrence from wake to deep sleep, and hence, it can potentially be exploited to detect sleep stages and/or sleep transitions. As discussed in Section 2.1.1 muscle activity can be measured by accelerometers.

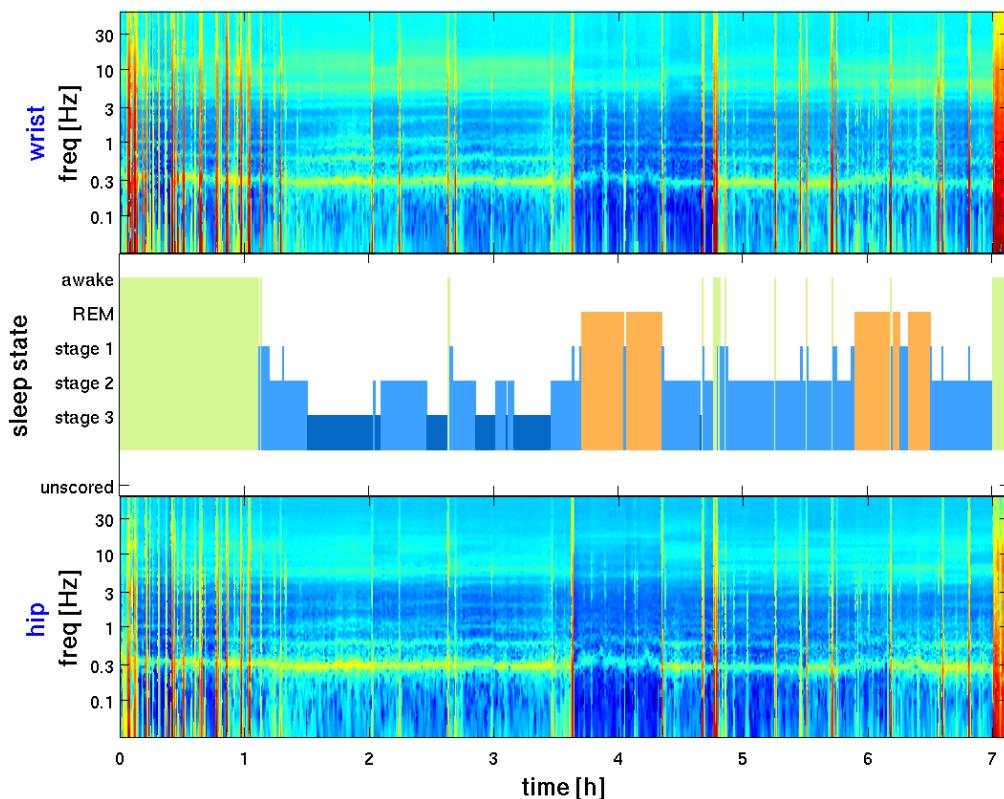


Fig. 3.1: Frequency spectrum of acceleration recordings of a whole night measured by SOMNOWatch. The upper and lower spectrograms show the frequency spectrum of accelerations at a subject's wrist and hip. In the middle the subject's hypnogram based on PSG is shown (awake coloured green, REM orange, light sleep light blue and deep sleep dark blue)

Figure 3.1 shows a typical frequency spectrum of accelerations for a whole night and the corresponding hypnogram based on polysomnography. For calculating the frequency spectrum the measured time signals in x-, y- and z-direction are divided into windows of 30 seconds duration. The beginning and the end of the recordings are given by the time period of the polysomnographic classification. The frequency spectrum of every window is obtained by an FFT. The time-dependent frequency spectrum is the consequence of

the vector sum computation of the frequency spectrum in x-, y- and z-direction of every 30-second window followed by a logarithmic binning of every timeframe. In Figure 3.1 the frequency spectrum is depicted as a spectrogram with a logarithmic frequency axis where the power of a given frequency is color coded. In this case red means a high, green a medium, and blue a low occurrence of a specific frequency. A qualitative relation between frequency f and intensity I can be understood by means of a calibration curve (see Figure 3.2). Note that intensities are multiplied by frequency f before colour coding. This procedure is chosen to compensate for the $1/f$ noise in the original data.

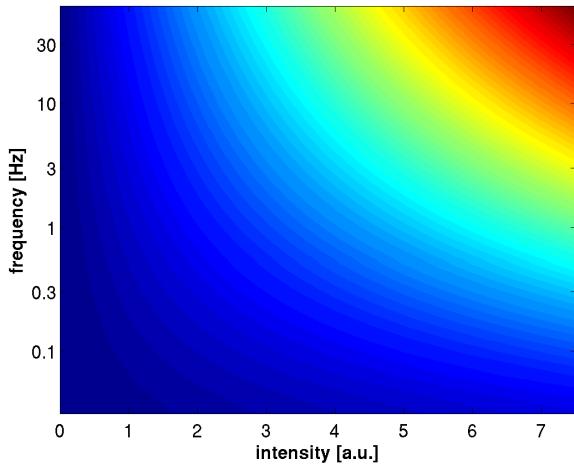


Fig. 3.2: Calibration curve for colour coding of power spectra. For different spectral intensities (horizontal axis) and different frequencies (vertical axis) a logarithmic colour coding is chosen from low intensity (blue) to high intensity (red)

Figure 3.1 shows stronger movements without a preferred frequency during wake compared with NREM sleep and REM sleep, respectively (red vertical lines in frequency spectrum). The physiology during REM sleep is characterised by a strongly reduced muscle tone and a shallow breathing. This is reflected by a low spectral power in both accelerometric signals from the hip and the wrist, respectively (see Figure 3.1 REM sleep episode at approximately 3.6h - 4.4h). However, one observes strong variations among different REM episodes in a single subject and among different subjects (see REM sleep at 6h, a high portion is coloured light blue and some areas are coloured green or yellow). Falsely reduced spectral power might be originated in probands lying on the accelerometer yielding an unwanted distortion of the recording.

When examining Figure 3.1 it is apparent that almost every transition from deep sleep to light sleep or wake is associated with a strong body movement. Arm movements or roll overs in bed are typical movements during a sleep stage transition which are indicated by high intensities in the frequency spectrum without a preferred frequency (yellow or red vertical lines in frequency spectrum, for example at 2h, 2h40min, 3h40min, 4h45min and 5h30min). High intensities in the frequency band from 0.2 Hz - 0.3 Hz are observed across all sleep stages (coloured yellow or green), however, they are reduced during REM sleep due to an overall reduced spectral power caused by a sizeable reduction in muscle tone (coloured light blue). A frequency of 0.2 Hz - 0.3 Hz translates to a period of 5 to 3.3 seconds; it corresponds to human respiration during rest with an average frequency of 11 - 15 breaths per minute (adults). Additional periodic signals with a high intensity (but less than the intensity of human respiration) and a frequency equal to an integer multiple of the respiratory base frequency are apparent (horizontal green or light blue lines, for example at roughly 0.6 Hz and 0.9 Hz - 1 Hz). These phenomena show the harmonics of breathing caused by the fact that human respiration cannot in general be described by a simple sine function of fixed base frequency f , but is rather a linear combination of that first harmonic and its (possibly infinitely many) higher harmonics of multiple integer frequencies ($2f, 3f, \dots$).

Another frequency band with a high intensity ranging from 6 Hz to 10 Hz is also in contrast to its surrounding frequencies (marked green in almost all sleep stages). This periodic signal, which does not occur in all subjects, exhibits a wider frequency spectrum as the human respiration. However, currently the signal cannot be identified unambiguously as a specific physiological phenomenon. Possibly, it is associated with lung sounds caused by lung rattling or snoring.

3.1.2 Power Spectrum Characteristics

After obtaining a typical frequency spectrum of a nocturnal accelerometric measurement, the basic findings must be scientifically examined. Therefore the calculation of a power spectrum of acceleration recordings is used. The power spectrum is a positive real function which describes the dependence of a movement intensity I on the movement frequency f of a human (see Section 2.3.2).

For calculating the power spectrum the measured time signal is divided into windows of 30 seconds duration each. The beginning and the end of the recording is determined by the duration of the polysomnographic classification. The frequency spectrum of every window is obtained by an FFT. Subsequently, all spectra are arithmetically averaged with respect to sleep stages (NREM sleep S1, S2, and S3, as well as REM sleep) and wake (S0). Finally, the five different averaged frequency spectra are binned logarithmically and plotted in a double logarithmic coordinate system where straight lines indicate a power-law behaviour (see Figure 3.3).

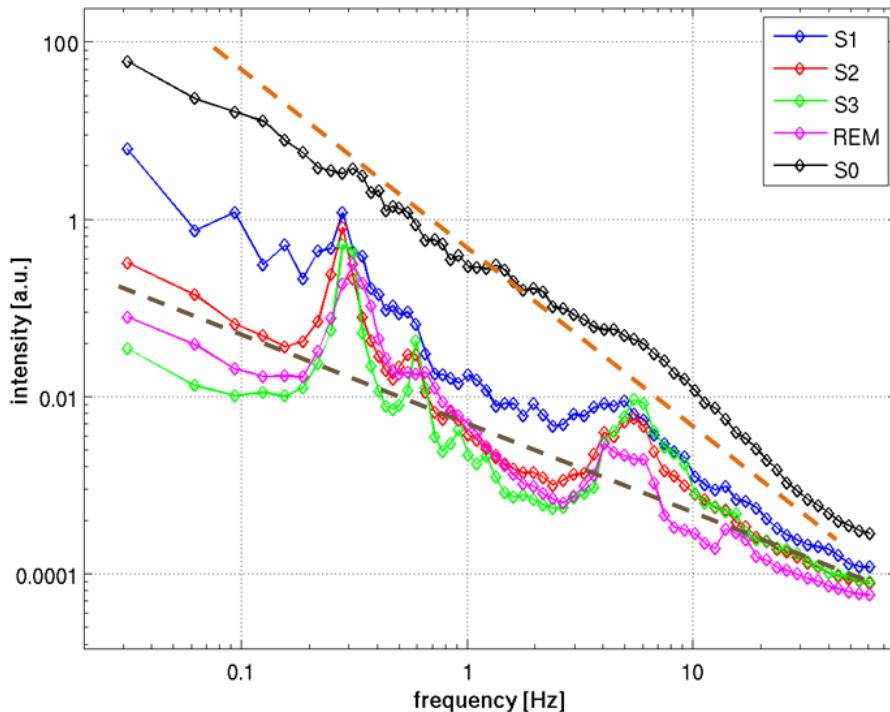


Fig. 3.3: Power spectrum (derived by FFT) calculated from raw SOMNOwatch data of a nocturnal accelerometric recording (healthy subject in a sleep laboratory). The data for different sleep stages (NREM sleep S1, S2, and S3, as well as REM sleep, colours) and wake states (S0, black) are separated. The dashed lines with slopes -1.0 and -2.0 are shown for comparison.

Figure 3.3 shows a typical double logarithmic power spectrum of a nocturnal accelerometric recording measured by SOMNOwatch. The represented curves exhibit in general a similar trend, i.e., an increasing frequency is associated with a reduced intensity. However, the coloured curves (different sleep stages) differ strongly from the power-law trend at two positions whereas the awake spectrum is closer to a power law.

The curves of the different sleep stages show two distinct maxima. The narrow peak in the frequency range from approximately 0.2 Hz to 0.4 Hz corresponds to the human respiration as discussed previously for spectrograms (see Section 3.1.1). The power in the respiratory frequency band is by a factor of 100 larger when compared with the background (see S3 sleep in green). The respiratory background is calculated by averaging the power in the frequency range from 0.1 Hz to 0.4 Hz. There are three additional peaks with decreasing amplitude at frequencies of 0.6 Hz, 0.9 Hz, and 1.2 Hz which are integer multiples of the respiratory base frequency of approximately 0.3 Hz - corresponding to above discussed higher harmonics.

The wide peak in the frequency range from 4 Hz to 10 Hz (see for instance S3 sleep) corresponds to a specific physiological phenomenon which is possibly associated with lung sounds caused by snoring or lung rattling. This maximum is characterised by a factor of 16 larger spectral power when compared with the baseline power (computed by power-law regression at the frequency of the power maximum), however, additional studies are needed for a more comprehensive physiological interpretation.

The two dominant peaks during sleep are not observed for the wake stage due to more frequent and intense movements during wake. The spectral power for wake is larger than the spectral power during sleep for all frequencies and for all sleep stages. Hence, respiratory motions are superimposed by stronger movements, such as arm or leg movements, during waking states. Similar results have been found in the spectrograms.

Excluding the two dominant peaks, i.e., the frequency ranges from 0.2 Hz to 0.4 Hz and from 4 Hz to 10 Hz, enables a power-law regression for all five power spectra. The effective spectral exponents obtained from the fits can be used to distinguish wake and sleep stages. The fitting function of the awake spectrum (determined by linear regression in the double logarithmic regime) has a slope close to -2 compared with a slope close to -1 for the sleep spectra. A slope of a linear function in a double logarithmic power spectrum with a value of -1 translates into a physical relationship of $I(f) \propto f^{-1}$. Random $1/f$ behaviour is found in many natural systems and generally referred to as $1/f$ noise. Many physiological signals, including heartbeat or neural activity, exhibit $1/f$ behaviour [13–17]. Time series analysis of long-term (24 hours) electrocardiograms (ECG) has shown $1/f$ characteristics in human heart rate and beat-to-beat (RR) interval fluctuations; for a first article on this subject from 1982 see Kobayashi and Musha [14]. Takahashi *et al.* determined $1/f$ -noise dynamics in firing (emitting action potentials) ventroposterior neurons during REM sleep over a range from 0.04 Hz to 1.0 Hz [15]. The production of a given spatial and temporal interval contains errors which fluctuate especially as $1/f$ noise [16]. Haussdorff *et al.* researched complex fluctuations in the normal gait pattern of ten healthy young men as they walked for nine minutes at their usual rate and tried to describe these fluctuations as being (i) uncorrelated white noise, (ii) short-range correlations, or (iii) long-range correlations with power-law scaling. He could demonstrate that conventional models of gait generation fail and a new type of central pattern generator model works successfully [17].

The spectral exponent of -2 during wake corresponds to a physical relationship of $I(f) \propto f^{-2}$, which is sometimes denoted as "brown noise" as it is characteristic of Brownian motion (random walks). It is significantly different from the power-law observed during sleep. The spectral exponent can, consequently, be exploited to distinguish sleep states from wake states. However, the power-law behaviour during sleep does not significantly differ between the four different sleep stages. For that reason it is not possible to implement an automated sleep stage classification solely based on spectral exponents calculated from acceleration data.

3.1.3 Comparison between Day and Night

Subsequent to an extensive study of nocturnal accelerometric measurements in sleep laboratories all-day accelerometric recordings are examined. The availability of diurnal and nocturnal data from the same subjects allows for comparison of different human movement patterns during day and night. For this purpose the accelerometric recording is subdivided into windows of 30 seconds duration each. The further approach is the same as the one described in Section 3.1.1 except that here the whole recording (day and night) is used and in contrast to the analysis of purely somnographic data above.

Figure 3.4 shows an exemplary 24h spectrogram obtained from an accelerometric recording using SOMNOWatch at a healthy subject's wrist. Note the logarithmic frequency axis and

the choice of colours from dark blue indicating low spectral power to red indicating high spectral power (see calibration curve in Figure 3.2). The time 0 on the linear time axis corresponds to the beginning of the accelerometric measurement and not to actual time. The subject was sleeping in a sleep laboratory for the first eight hours of the accelerometric recording while he spent the remaining time following his typical daily routine.

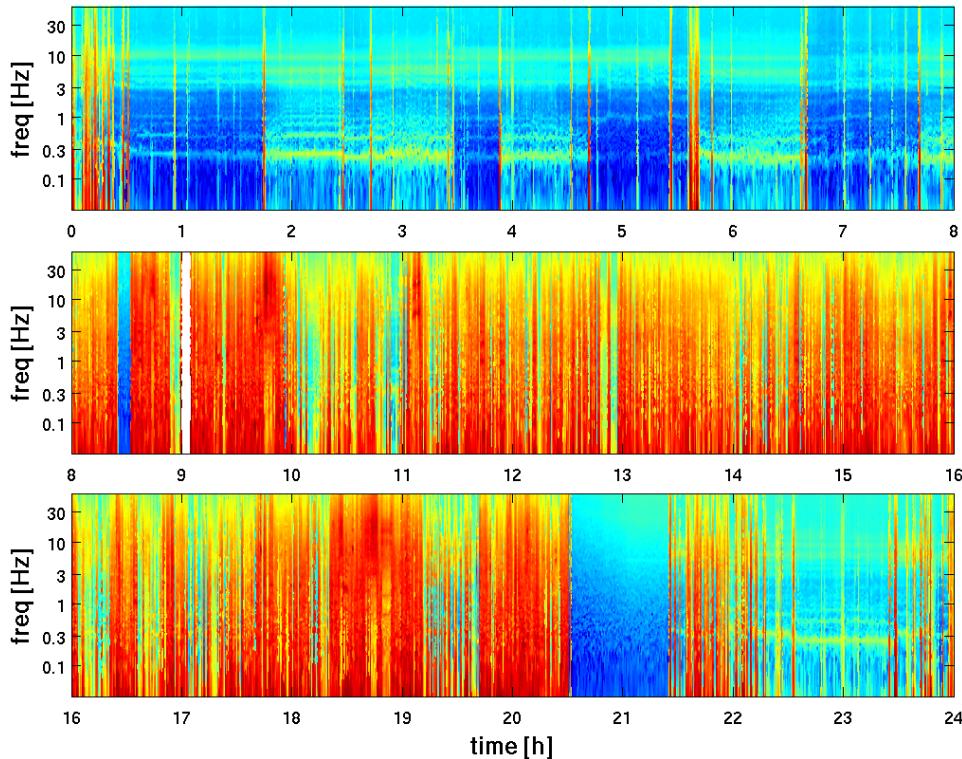


Fig. 3.4: 24h spectrogram calculated from an accelerometric recording in a healthy subject using SOM-NOWatch at the subject's wrist. Times 0h - 8h and 22h - 24h correspond to nocturnal data and times 8h - 22h correspond to diurnal data.

During daytime from 8h to 22h a strikingly increased intensity compared with sleep periods (hours: 0-8 and 22-24) can be observed. The high spectral power (red and yellow patches) during most of the day indicates a large amount of (voluntary) movements without a characteristic frequency. The elevated spectral powers in all frequencies due to body movements mask the respiratory oscillation and its higher harmonics that are expected to be present at integer multiples of the base frequency of 0.3 Hz. Using the characteristic features of movement-induced spectral power during daytime and night both physiological states can be well distinguished by spectral analysis of accelerometric data.

The one-hour period starting at 20h30min is characterised by low spectral intensity for all frequencies (blue colouring). It shows no definite structures such as respiration. The absence of movement activity suggests that the subject took off the accelerometric device and put it on a calm place. The acceleration measurement during this time is a nearly constant signal which varies only due to the measuring inaccuracy of the accelerometer. Hence, an FFT of a 30-second time window during that period results in an (almost) constant intensity I_c for all frequencies f . Indeed, the intensities shown in the spectrogram are not constant but increase with increasing frequency (dark blue to light blue). That is a result of the calibration curve whereby the colour is a measure of $f \cdot I_c$. The same feature can be seen in a roughly 8-minute time window around 8h30min.

In the spectrogram there is a white 5-minute segment after 9h. During this time the accelerometer records no real data but saves the minimal value of the measuring range. This is due to a programming error of the device software. An FFT of this constant function yields $-\infty$. That is the reason for the white time window in the spectrogram.

3.2 Comparison of Accelerometric Devices

In the context of the German National Cohort pre-study I three different accelerometric devices - SOMNOwatch, Actigraph GT3Xplus and GENEactiv - are used for long-term (up to seven days) acceleration recordings. The technical data of the three accelerometers is described in Section 2.2. After testing and comparing all three devices the German National Cohort has to select one accelerometer for the main study with 200,000 subjects.

For comparing the functionality of all three accelerometers representative 10-seconds time windows with very weak movement activity during quite sleep are chosen (see Figure 3.5). The representative timeframes show only accelerations along one axis and are selected from accelerometric recordings measured at the subject's hip. During quiet sleep body motions are mainly induced by respiration. In Figure 3.5 respiratory activity is evident with a rate of approximately three breaths per 10 seconds and peak-to-peak amplitudes of about $0.02g$. But only the records from SOMNOwatch and Actigraph GTX3plus show the typical behaviour, whereas the GENEactiv signal is disturbed by a strong measurement noise.

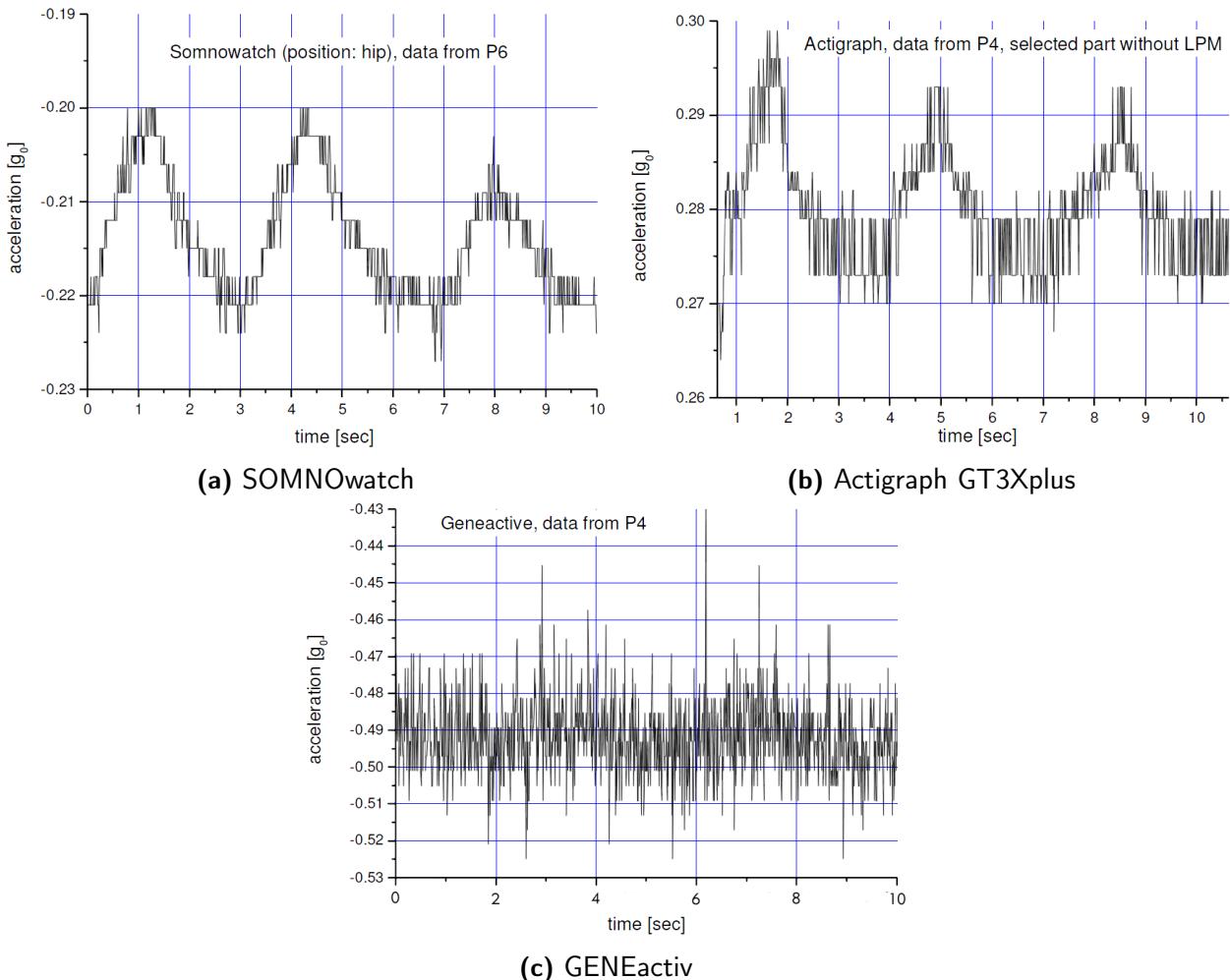


Fig. 3.5: Accelerations at a subject's hip measured by three different accelerometers during a 10-second episode of quiet sleep with very weak activity. The depicted timeframes are selected from accelerometric recordings of the P6 study (SOMNOwatch, Actigraph GT3Xplus) and P4 study (GENEactiv).

Figure 3.5 also emphasizes two minor deficiencies of the Actigraph GTX3plus device. In addition to a slightly higher noise level compared with SOMNOwatch the digital resolution intervals are not identical. This limitation becomes obvious for, e.g., accelerations between $0.273g$ and $0.279g$ which cannot be resolved (resolution interval $0.006g$), while acceleration values of $0.282g$ and $0.284g$ are recorded (resolution interval only $0.002g$). These considerable differences of possible resolution intervals (variation by a factor of three) are generally

untypical for a measuring instrument. Accelerometric recordings of SOMNOWatch do not exhibit such non-equidistant resolution intervals, i.e., all resolution intervals are close to $0.003g$.

In order to determine whether GENEactiv recordings are generally disturbed by high noise levels 7-day data are studied in the following. The raw acceleration data is subdivided into 30-second segments and only segments with very low motion activity are selected for further analysis by using a standard deviation criterion. For this purpose the standard deviation of the measured accelerations of each 30-second time window is calculated and if a standard deviation is greater than a specific threshold, this segment is skipped due to a high motion activity. Figure 3.6 shows the averaged power spectra calculated from selected low-motion segments for six different subjects on double-logarithmic axes.

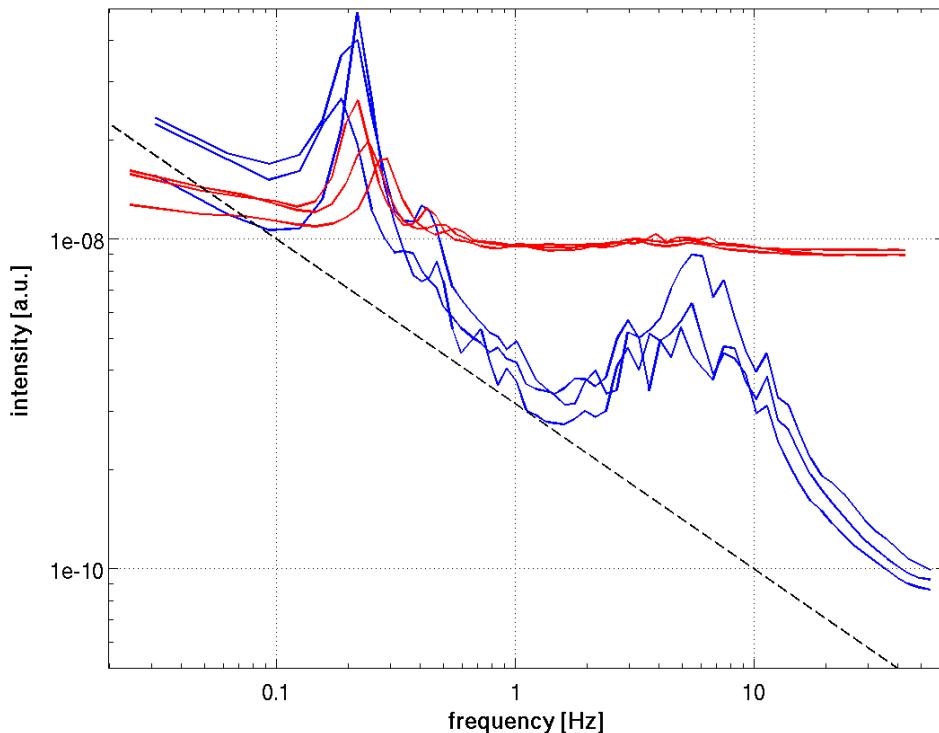


Fig. 3.6: Power spectra of accelerometric recordings measured by SOMNOWatch (blue, data from P6 study) and GENEactiv (red, data from P4 study) in six different subjects. The raw acceleration data (divided in 30-second segments) with very low motion activity was selected by setting a threshold for the corresponding standard deviation.

The strong noise in GENEactiv recordings (red) is also reflected in Figure 3.6. With the aid of these power spectra the consequences of the high noise level can be studied and the spectral features of a long-term GENEactiv measurement can be characterised. Though Figure 3.5 suggested, that even the strong human respiratory oscillation might not be clearly be detected by GENEactiv due to noise, a distinct narrow peak in the frequency range from approximately 0.2 Hz to 0.4 Hz is clearly recognisable. However, all spectral properties above 0.5 Hz are lost because of the underlying noise. The flat power spectrum with an intensity level corresponding to 10^{-8} in the double logarithmic plot suggests that the underlying noise is white noise, characterized by a zero spectral exponent. It is most probably an intrinsic problem of the measuring sensor.

Note that it can be confirmed that high noise levels only affect episodes of low motion activity. Stronger movements result in a higher intensity and consequently the sensor noise has a reduced influence. Since acceleration measurements during sleep correspond to times of reduced body movements a detection and classification of sleep is severely limited by the strong white noise in the GENEactiv accelerometer.

In contrast to the GENEactiv device the SOMNOwatch accelerometer has a strikingly decreased noise level (at least by two orders of magnitude, see blue lines in Figure 3.6) and is, hence, better suited for sleep detection.

Subsequently to an evaluation of the GENEactiv device the characteristics of Actigraph GT3Xplus are tested in the following. Figure 3.7 shows a comparison of accelerations measured at a subject's hip by SOMNOwatch (left) and Actigraph GT3Xplus (right) during the same motion event. The SOMNOwatch recording clearly shows the human respiration as a periodic signal starting at a time of about 1557 seconds (especially in x- and y-direction). At the same position in the Actigraph GT3Xplus recording (after time 1724s) there is no continuous periodic signal, but a spike train of 1s period. Therefore the respiratory activity, which could be recorded at the given accuracy (see 1715s till 1724s), is not correctly measured. The different time designations of the motion event are due to a different beginning of the two accelerometric recordings. The comparison with the SOMNOwatch signal further indicates that Actigraph GT3Xplus misses the initial part of the motion activity (signals begin with a jump at 1685s for Actigraph GT3Xplus, but they begin continuously at the corresponding 1525s for SOMNOwatch). This is due to a "feature" called "low power mode" (LPM) implemented by the Actigraph company. According to the device's manual, the sampling rate is reduced from 80 Hz to 1 Hz if a nominal activity below a predefined threshold is detected for more than 10 seconds. The accelerometer continues recording at 80 Hz only if motion of a specific strength in x- or y-direction is detected in one of the measurements taken at 1 Hz, which is not representative for the full preceding second.

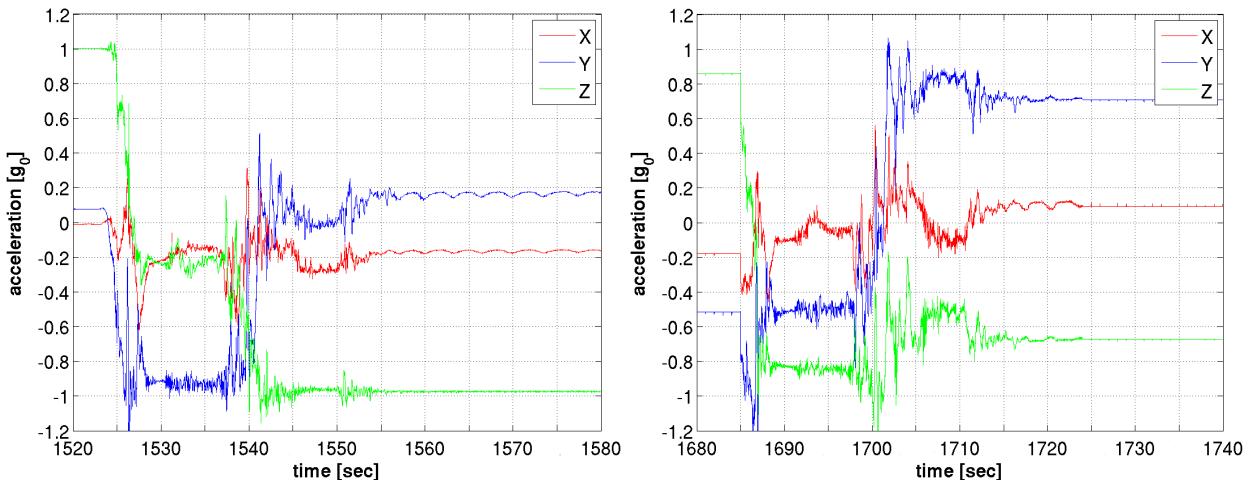


Fig. 3.7: Accelerations at a subject's hip measured by SOMNOwatch (left) and Actigraph GT3Xplus (right) during the same motion event (data from P6 study).

The statistical analysis of nocturnal Actigraph GT3Xplus recordings of five different subjects in a sleep laboratory shows that the accelerometer suspends into LPM during approximately 94 percent of nighttime. A following study of six different 7-day Actigraph GT3Xplus recordings shows that more than 55 percent of all data (day and night) are effectively recorded at only 1 Hz. On the basis of these two pieces of information and under the assumption that a human sleeps on average 7.5 hours per day the total percentage of LPM during daytime can be calculated (see Table 3.1).

Tab. 3.1: Statistical analysis of six 7-day Actigraph GT3Xplus recordings from study. The total percentage of measured data in LPM is derived for each subject, and the percentages of LPM during daytime are approximated assuming 7.5 hours of sleep per day and 94 percent of LPM during nighttime

Subject no.	1	2	3	4	5	6	average
Total LPM [%]	52.8	59.2	53.5	55.3	57.5	55.7	55.7
Daytime LPM [%]	38.8	49.3	39.9	42.9	46.6	43.6	43.6

If the accelerometer is in LPM, the measured data of one second consists of 79 constant dummy values and only one real measured value. In fact, the dummy value is equal to the acceleration value that was measured when the device suspended into LPM; possibly a longer time, ago. Since only one out of 80 values is meaningful and different from constant, data during LPM looks like a spike train (see Figure 3.8). The device resets to the normal measurement mode if a spike exceeds a specific threshold. But the thresholds for switching into LPM and back are proprietary information and not given in the device's specifications. They can only be estimated from the spikes' magnitudes.

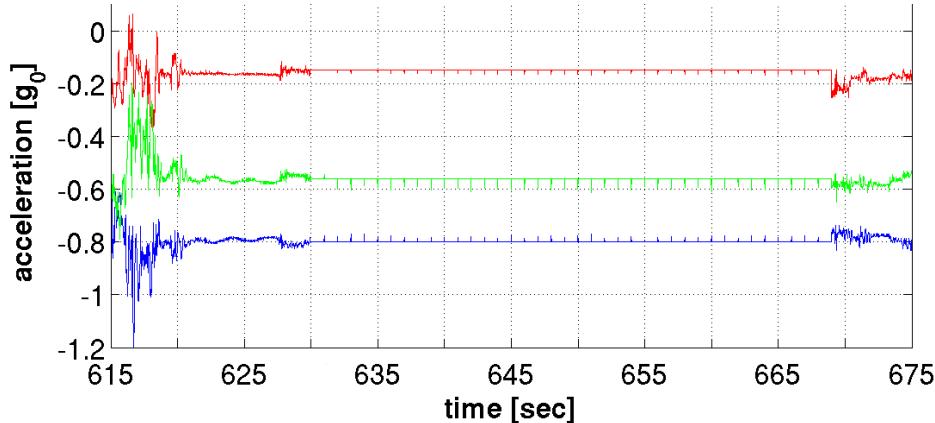


Fig. 3.8: Accelerations at a subject's hip measured by Actigraph GT3Xplus. Most of the depicted signal (from 630s to 669s) is spent in Low Power Mode (LPM), i.e., the measured data mostly consists of dummy values with spikes in one-second intervals.

The spikes in LPM and the permanent alternation of normal mode and LPM hinders any data evaluation, analysis and related derivations (see, e.g., the comb-like artefacts in the semi-logarithmic power spectrum in Figure 3.9 which are only caused by LPM).

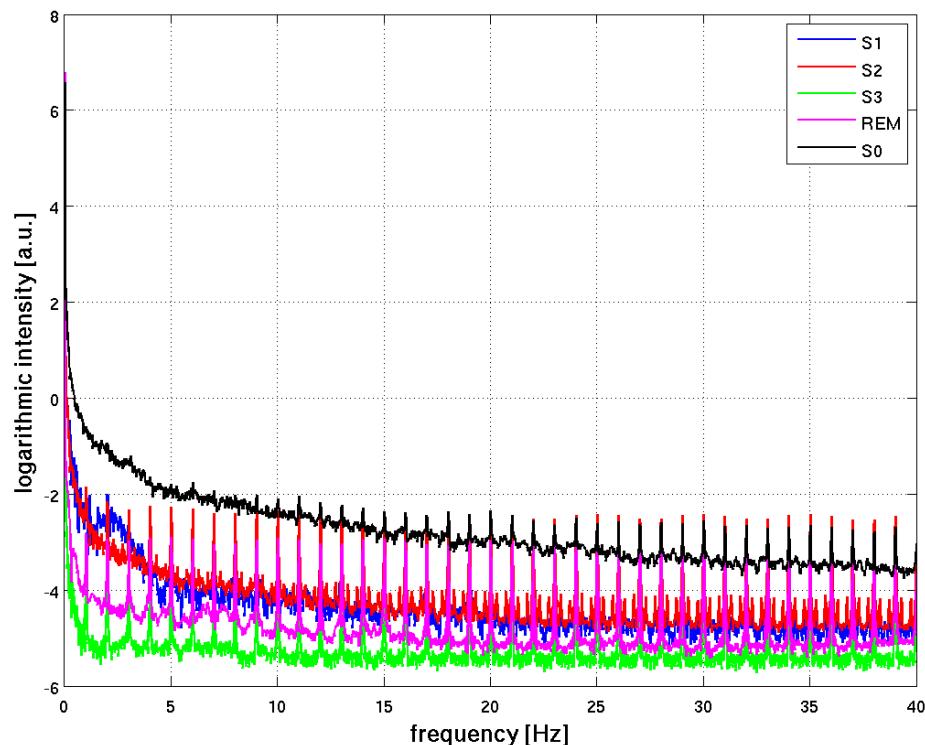


Fig. 3.9: Semi-logarithmic power spectrum (derived by FFT) of an exemplary nocturnal accelerometric recording measured by Actigraph GT3Xplus. The data for different sleep stages (NREM sleep S1, S2, and S3, as well as REM sleep, colours) and wake states (S0, black) were separated prior the analysis. The comb-like spikes are artefacts due to LPM. They spoil significant parts of the spectrum even during wake states.

The results suggest that Actigraph GT3Xplus is not suited for sleep studies aimed at detecting sleep onset and sleep stage transitions. According to the device's manual and further details from the company's support team the LPM cannot be disabled. The characteristics of the LPM, including a reduced sampling rate of only 1 Hz during roughly 14 hours a day and especially during more than 95 percent of the total sleep time, questions the implementation of established sleep-aware-differentiation algorithms as developed by Cole and Kripke in 1992 [1] or by Sadeh and Sharkey in 1994 [2]. Both algorithms were specifically designed for time-continuous accelerometer recordings with a constant sampling rate (greater than 1 Hz) in order to resolve all motion activity. Since the algorithms are still implemented in the device's software (without noting possible limitations) doubts about the device's reliability and the developer's trustworthiness arise. In conclusion, the Actigraph GT3Xplus device with the current LPM implementation cannot be utilised to develop an advanced sleep-aware-differentiation algorithm.

After consulting the manufacturer of the Actigraph GT3Xplus accelerometer they were determined to develop a firmware update for the device. At the request of the German National Cohort this update should eliminate the bad features of the LPM or provide the opportunity to deactivate the LPM. Within a short time the firmware update was completed and 7-day recordings with the revised Actigraph GT3Xplus were realised. The sampling rate was raised from 80 Hz to 100 Hz.

Figure 3.10 shows accelerations at a subject's hip measured by the revised Actigraph GT3X+ during sleep (first night of a 7-day recording). It is apparent that a revised LPM (here called LPM2) was implemented. The Actigraph GT3Xplus is in LPM2 between 127s and 128s and again between 139s and 154s (see right plot of Figure 3.10). During LPM2 the measured signal is constant, i.e., only dummy values (measured value at the time of suspension into LPM2) are saved. They do not contain any information. Suspension into LPM only starts or ends at full seconds whereby the initial part of a motion activity might be missed.

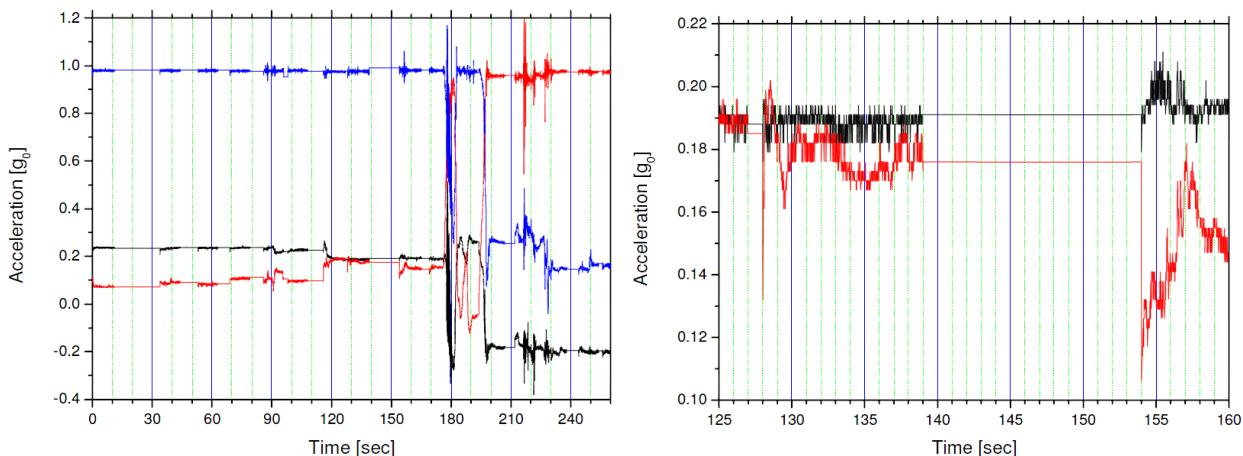


Fig. 3.10: Accelerations at a subject's hip measured by the revised Actigraph GT3Xplus during sleep (data from P4 study, all three axes x, y, z shown in different colours). The right plot is a magnification of the left plot.

The statistical analysis of three different 7-day accelerometric recordings measured by the revised Actigraph GT3Xplus shows that the device suspends into LPM2 roughly during 23 percent of the total recording time which is an improvement of 20 percent over LPM, though, about a quarter of the data is still unusable. The thresholds used for switching between normal operation and LPM remain undisclosed information. Due to the lack of spike trains during LPM2 an estimation of the thresholds is no longer feasible. However, the threshold for returning into normal measuring mode seems to be significantly lower (compared with LPM), resulting in a reduced overall time spent in LPM2. Missing transients during

LPM2 have the positive side effect of simpler data processing since no additional methods are necessary to obtain less artifact-disturbed results when applying FFT. In conclusion, this firmware update has some advantages, although, not all requested modifications were implemented.

Figure 3.11 compares typical power spectra from accelerometric recordings measured by SOMNOwatch, GENEactiv, and Actigraph GT3Xplus (after firmware update) in six different subjects. It is obvious that (i) the revised Actigraph GT3Xplus can now be used for developing a spectral analysis based sleep-aware differentiation algorithm and that (ii) its noise level is clearly intermediate compared with the other two accelerometers, making the Actigraph GT3Xplus device clearly superior to the GENEactiv accelerometer. Nevertheless, the SOMNOwatch device remains the best suited device for the goals of the German National Cohort study, because it is not affected by the disadvantages of LPM/LPM2 or a similar power saving mode and it is characterised by an overall lower noise level.

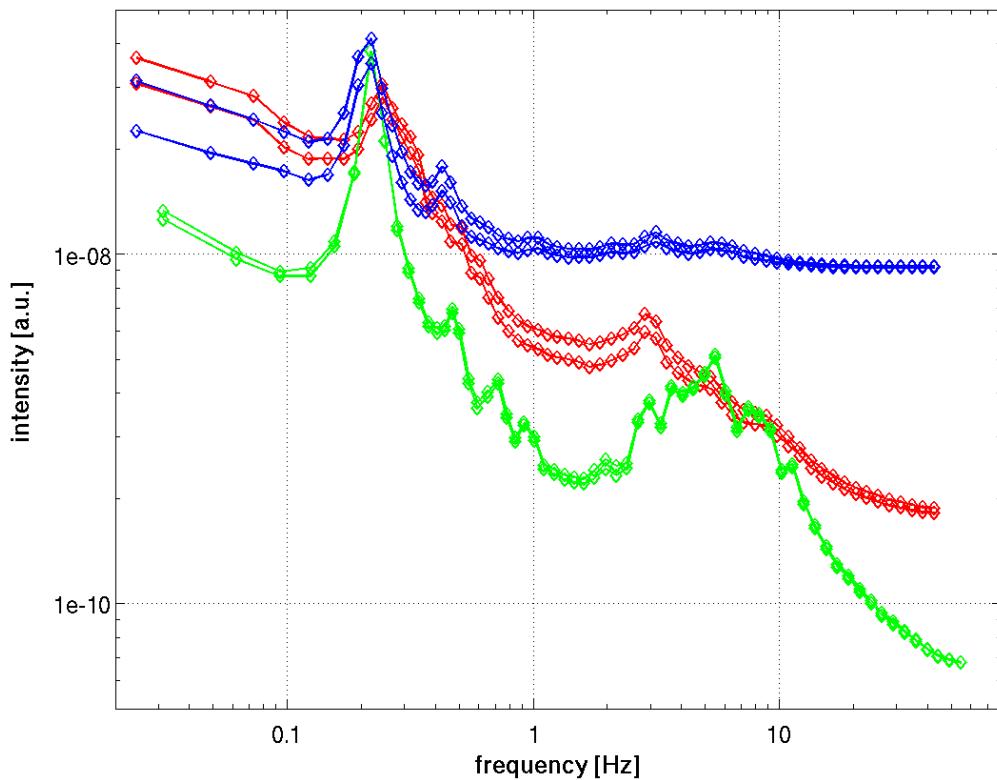


Fig. 3.11: Power spectra of accelerometric recordings measured by SOMNOwatch (green, data from P6 study), GENEactiv (blue, data from P4 study), and Actigraph GT3Xplus after firmware update (red, data from P4 study) in six different subjects. The raw acceleration data (divided into 30-second segments) with very low motion activity was selected by setting a threshold for the corresponding standard deviation.

If the Actigraph GT3Xplus device is used in the main study of the National Cohort comprising 200,000 subjects, a further firmware update will be needed, because the data losses due to LPM/LPM2 are not acceptable.

3.3 Development of a Sleep-Awake Differentiation Algorithm

For developing a sleep-aware differentiation (SAD) algorithm based on acceleration recordings the polysomnographic sleep classification serves as a reference. During the P6 study in the context of the German National Cohort nocturnal accelerometric data of 100 subjects were recorded in sleep laboratories using SOMNOwatch devices at the subjects' wrist and hip as well as using Actigraph GT3Xplus at the subjects' hip. Since the measurements were obtained before the firmware update of Actigraph GT3Xplus, its data cannot be used to distinguish awake and sleep. Therefore a SAD algorithm can only be developed for the SOMNOwatch accelerometer. For this purpose 30 subjects (out of 100) are randomly chosen to form the training sample. On the basis of the remaining 70 subjects (control sample) the newly developed algorithm is tested.

3.3.1 Parameter Selection

In this thesis the main method for developing a SAD algorithm is the spectral analysis of accelerometric recordings. To ensure better comparability with the polysomnographic sleep classification, which is done in 30s intervals, the measured time signal is divided into windows of the same length. Afterwards the frequency spectrum of every window is obtained by an FFT and the power spectra are averaged with respect to sleep stages and wake. The averaged spectra are depicted on double logarithmic axes in Figure 3.12.

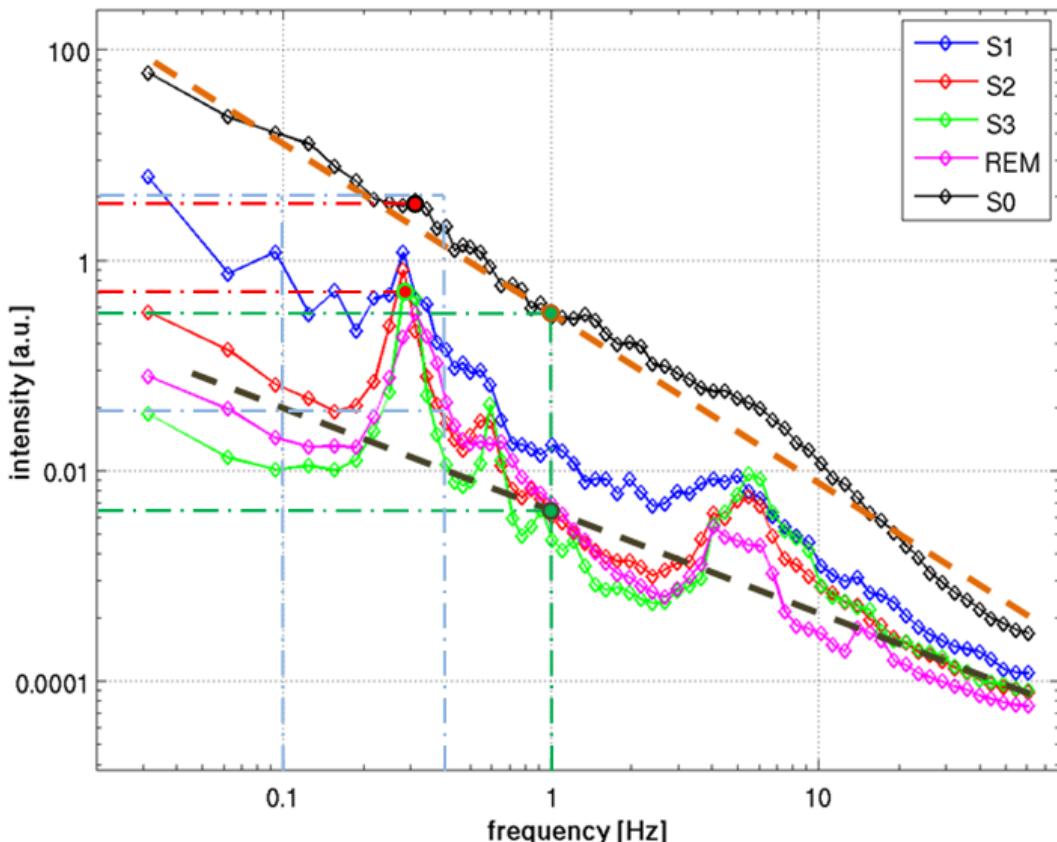


Fig. 3.12: Averaged power spectra from nocturnal accelerometric data recorded with SOMNOwatch (healthy subject). Different colours indicate different sleep stages (NREM sleep S1, S2, and S3, as well as REM sleep) and wake states (S0). The dashed lines are fits to the data for wake (orange line) and REM sleep (brown line).

To reliably distinguish sleep from wake variables (scores) that shall take a value above a cutoff for one state and below the cutoff for the other state are needed. Here, the relevance of three parameters calculated from power spectra is tested.

Neglecting the two dominant peaks from respiration and snoring between 0.2 Hz - 0.4 Hz and 4 Hz - 10 Hz, respectively, the power spectra approximately follow a power law. Therefore, the data is best fitted on double logarithmic axes by implementing a linear regression (see Figure 3.12, coloured dashed lines). The slope s and the intercept i of the linear relation $y = s \cdot x + i$ are already two candidate variables whose predictive power is worth checking. The slope reflects the dependence of intensity E on frequency f and in a range of roughly 30 mHz to 70 Hz and the intercept is a measure of the intensity at a specific frequency, i.e., it describes the power of motion activity. In Section 3.1.2 it was shown, that the slope of wake states is close to -2 and the slope during the different sleep stages is close to -1 . In addition, the frequency analysis reveals a higher amount of powerful body movements during wake phases than during sleep. Consequently, the slope and the intercept are distinctive features of sleep and awake.

In this work the intercept i of the linear fit is defined at the logarithmically averaged frequency f of N binned data points with frequency f_n and averaged intensity E_n :

$$i = \langle \log_{10} E \rangle - s \cdot \langle \log_{10} f \rangle \quad \text{where} \quad \langle \log_{10} x \rangle = \frac{1}{N} \sum_{n=1}^N \log_{10} x_n \quad (12)$$

The slope s follows directly from the linear regression:

$$s = \frac{\sum_{n=1}^N (\log_{10} f_n \cdot \log_{10} E_n) - N \cdot \langle \log_{10} f \rangle \cdot \langle \log_{10} E \rangle}{\sum_{n=1}^N (\log_{10} f_n)^2 - N \cdot \langle \log_{10} f \rangle^2} \quad (13)$$

The green dash-dotted lines in Figure 3.12 show the intensity at the frequency $f = 1$ Hz. As can be seen the intensity during wake states is about 90 times higher than the intensity during REM sleep.

The third parameter is derived from the respiration peak. Since the sleep stages have less motion activity the respiration is better detectable during sleep than during wake. The idea is to use the difference, r , between the respiratory peak-power and the baseline power in the log-log power spectrum as a third parameter; consult blue (baseline) and red (respiratory peak) dash-dotted lines in Figure 3.12 for illustration. The respiratory background is calculated by averaging the logarithmic intensity in the frequency range from 0.1 Hz to 0.4 Hz. It is apparent that the relative height in linear scale during sleep is considerably larger than 1 (≈ 12 , S3 sleep) and during wake it is approximately 1 or smaller (≈ 0.8 , S0).

It appears that the shape of the possibly snoring and lung rattling related peak between approximately 4 Hz and 10 Hz strongly varies among different subjects, hence, it cannot be exploited as an additional sleep-wake predictor.

3.3.2 Classification Function

Above discussed parameters, slope s , intercept i , and respiratory peak-ratio r can be used to construct a score to distinguish sleep from wake. Therefore, s , i , and r are calculated for every (non-overlapping) 30-second time interval. The dynamics in time is depicted in the center panels of Figure 3.13. The upper central panel illustrates the determined variables of each 30-second interval without any processing. It is apparent that this temporal progress clearly fluctuates so that too much sleep/wake transitions would be determined for a natural human sleep behaviour. The implementation of a specific smoothing procedure (shown in the lower central panel of Figure 3.13) reduces this effect and yields a useful temporal behaviour for applying a multi-dimensional linear regression.

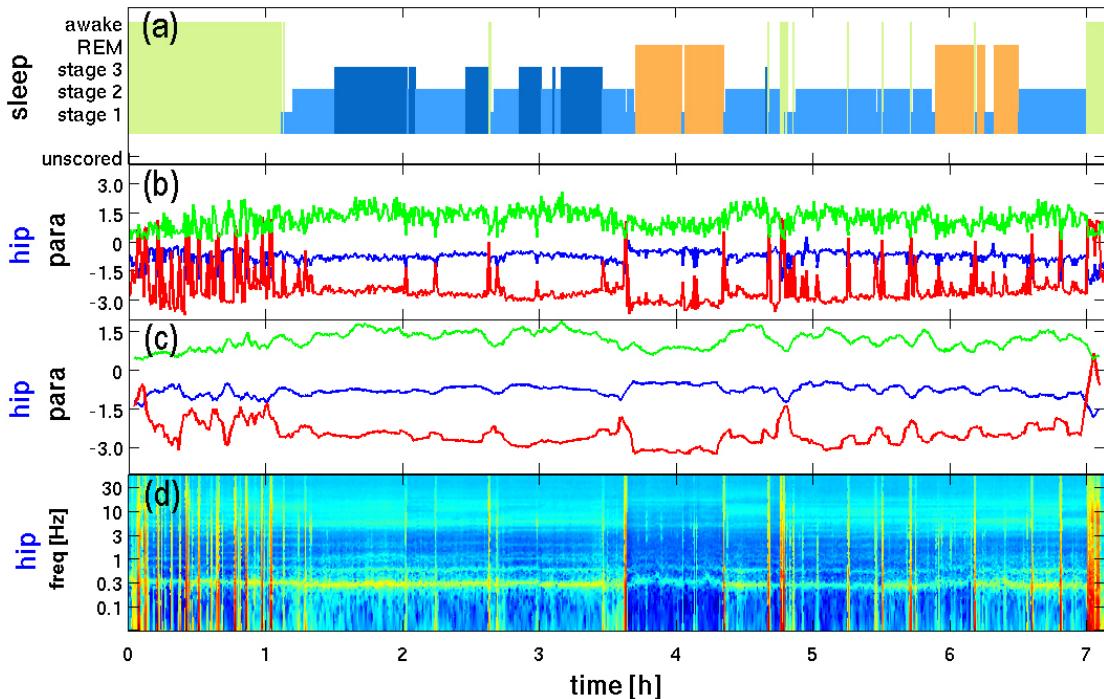


Fig. 3.13: (a) Hypnogram from a healthy subject obtained from full-night polysomnography. (b) Dynamics of power-spectral slope s (red), intercept i (blue), and respiratory peak-ratio r (green) calculated in 30-second intervals from accelerometric recordings using SOMNOwatch. (c) Same as (b) but smoothed by SMA ($\tau = 11$, see Section 2.3.3). (d) Corresponding spectrogram of the accelerometric raw data.

The optimal choice of both, the smoothing procedure and the linear combination, is found by comparing the sleep/wake prediction using the classification function λ with the polysomnographic sleep classification in a receiver operating characteristic (ROC). In that case, each value of the classification function λ greater than a specific threshold means a 30-second wake phase and all the other values imply a 30-second sleep state.

$$\lambda = \alpha \cdot i_{\tau}^p + \beta \cdot s_{\tau}^p + \gamma \cdot r_{\tau}^p + \delta \cdot i_{\nu}^q + \varepsilon \cdot s_{\nu}^q + \zeta \cdot r_{\nu}^q \quad (14)$$

where $\alpha, \beta, \gamma, \delta, \varepsilon, \zeta$... constant coefficients
 p, q ... smoothing procedures (see Section 2.3.3)
 τ, ν ... range of smoothing (see Section 2.3.3)

The coefficients in the classification function λ are devised in several optimisation steps. In general, an iteration consists of the choice of a smoothing procedure (SMA or EMA), the variation of τ or ν as well as the consideration of all possible linear combinations (variation of coefficients) of two independent variables at a constant τ or ν . For each linear combination a ROC curve is determined by moving the differentiation threshold. Subsequently, the area under the ROC curve (AUC) is calculated. Since the AUC is a measure of the concordance between the results of the classification function and the polysomnography, this value is optimised here for the linear combination in Equation (14). I.e., the optimal classification function exhibits the maximal AUC.

The first iteration comprises the optimisation of the first three terms of the linear combination λ (neglecting the last three terms). Without loss of generality, the coefficient α is set to one and the other two coefficients are subsequently optimised in the geneal iterative procedure. During the first iteration the two smoothing procedures SMA and backward EMA are used, the oneside timeframe τ is varied from 1 to 20 and the coefficients β and γ are optimised in the range $[-2.5, 2.0]$. After completing this iteration and comparing both smoothing approaches with each other, it becomes apparent that the backward EMA (bEMA) yields

better results than the SMA. That was already expected, because future events should not influence the present and the impact of past events on the present should decrease with time (memory). The obtained linear combination differs according to the measuring site (hip or wrist).

$$\lambda_{\text{hip}} = i_{11}^{\text{bEMA}} + 1.10 \cdot s_{11}^{\text{bEMA}} - 0.38 \cdot r_{11}^{\text{bEMA}} + \delta \cdot i_{\nu}^q + \varepsilon \cdot s_{\nu}^q + \zeta \cdot r_{\nu}^q \quad (15)$$

$$\lambda_{\text{wrist}} = i_9^{\text{bEMA}} + 1.30 \cdot s_9^{\text{bEMA}} - 0.75 \cdot r_9^{\text{bEMA}} + \delta \cdot i_{\nu}^q + \varepsilon \cdot s_{\nu}^q + \zeta \cdot r_{\nu}^q \quad (16)$$

Due to the long-range backward EMA the undefined part of the classification function is optimised by applying a short-range peaked EMA (pEMA) and SMA ($\nu \in [0, 5]$). These smoothing procedures (described in Section 2.3.3) are chosen to include the direct environment of the present event containing significant information for differentiating sleep and wake. Furthermore, ζ is equal to 0, because strong fluctuations exist in the respiratory maximum for short smoothing ranges during sleep and wake, and hence, those rapid changes do not provide any further information. After the second iteration procedure the optimal classification functions for hip and wrist data read:

$$\lambda_{\text{hip}} = i_{11}^{\text{bEMA}} + 1.10 \cdot s_{11}^{\text{bEMA}} - 0.38 \cdot r_{11}^{\text{bEMA}} + 0.26 \cdot i_1^{\text{pEMA}} + 0.22 \cdot s_1^{\text{pEMA}} \quad (17)$$

$$\lambda_{\text{wrist}} = i_9^{\text{bEMA}} + 1.30 \cdot s_9^{\text{bEMA}} - 0.75 \cdot r_9^{\text{bEMA}} + 0.20 \cdot i_1^{\text{pEMA}} + 0.08 \cdot s_1^{\text{pEMA}} \quad (18)$$

In another approach slope s , intercept i , and respiratory peak-ratio r are normalised before optimising the linear coefficients. Such normalisation is expected to allow for a better comparability between subjects, and shall therefore, (hopefully) result in a more precise differentiation of sleep and wake. The parameter time series are normalised to zero mean and unit variance using the standard normal distribution estimators for average value μ_x and the variance σ_x^2 , respectively.

$$\mu_x = \frac{1}{N} \sum_{n=1}^N x_n \quad \sigma_x^2 = \frac{1}{N-1} \left(\sum_{n=1}^N x_n^2 - \frac{1}{N} \left(\sum_{n=1}^N x_n \right)^2 \right) \quad (19)$$

The values x_n are transferred to the normalised values y_n by using Z score transformation, where N is the number of 30-second intervals of the measurement time:

$$y_n = \frac{x_n - \mu_x}{\sigma_x} = \left(x_n - \frac{1}{N} \sum_{n=1}^N x_n \right) \cdot \sqrt{\frac{N \cdot (N-1)}{N \cdot \sum_{n=1}^N x_n^2 - \left(\sum_{n=1}^N x_n \right)^2}} \quad (20)$$

After applying the optimisation process on the basis of the normalised parameters, the ideal classification functions are

$$\underline{\lambda}_{\text{hip}} = \underline{i}_{11}^{\text{bEMA}} + 0.54 \cdot \underline{s}_{11}^{\text{bEMA}} - 0.23 \cdot \underline{r}_{11}^{\text{bEMA}} + 0.11 \cdot \underline{i}_1^{\text{pEMA}} + 0.04 \cdot \underline{s}_1^{\text{pEMA}} \quad (21)$$

$$\underline{\lambda}_{\text{wrist}} = \underline{i}_9^{\text{bEMA}} + 0.41 \cdot \underline{s}_9^{\text{bEMA}} - 0.28 \cdot \underline{r}_9^{\text{bEMA}} + 0.14 \cdot \underline{i}_2^{\text{pEMA}} + 0.04 \cdot \underline{s}_2^{\text{pEMA}}. \quad (22)$$

The four obtained classification functions are quite similar. A comparison of the two main parts of the linear combination shows that the part smoothed by the long-range backward EMA has the largest impact on the result. Additionally, the influence of slope and respiratory peak-ratio in the normalised classification function is reduced when compared with its unnormalised counterpart. One possible reason might be a higher variance of slope and respiratory peak-ratio compared with the variance of the intercept during nighttime (see Figure 3.13).

3.3.3 Algorithm Quality

For estimating the quality of the developed sleep-aware differentiation algorithm the obtained results are compared with the polysomnographic sleep classification. The classification function of the SAD algorithm was optimised until the maximal area under the receiver operating characteristic (AUC) was reached. The final ROC curves of the four determined linear combinations (described in Section 3.3.2) are plotted in Figure 3.14.

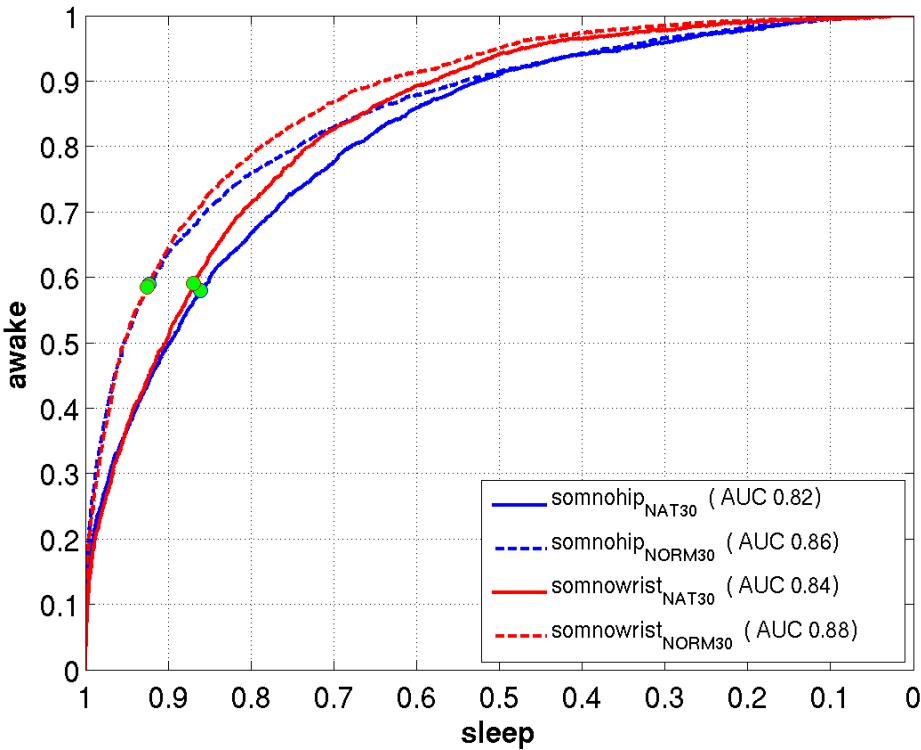


Fig. 3.14: Receiver operating characteristic of four different SAD algorithms (two measuring sites - hip and wrist, natural and normalised parameter series), where the correct wake scoring versus the correct sleep scoring (compared with the polysomnographic sleep classification) is illustrated. The final AUC values are reported in the legend. The four green dots mark the amount of correct sleep and wake scorings at the optimal threshold for calculating the total sleep time.

As can be seen each ROC curve exhibits an AUC in the range from 0.82 to 0.88. This is a very good result, because many currently established methods in medicine are characterised by an AUC smaller than 0.8. For instance, an AUC range from only 0.61 to 0.66 is obtained by comparing two screening tools (the Identification of Senior At Risk - ISAR, and the Triage Risk Stratification Tool - TRST) with the best well-known geriatric tool (the Comprehensive Geriatric Assessment - CGA) for identifying unplanned readmission of patients at risk. Nevertheless, these screening tools are medically used due to their low specificity and high negative predictive value [18].

The classification functions based on normalised data (see Figure 3.14, dashed lines) have the largest AUCs. Accordingly, the best results are achieved by accelerometric recordings at a subject's wrist with a subsequent parameter normalisation (red dashed line, AUC = 0.88). The differences in the results obtained separately from either hip or wrist measurements become apparent only if less than 90% of the sleep intervals are correctly identified (hip data's ROC runs below wrist data's ROC). For better sleep detection the two curves of the same data family (natural or normalised) behave quite similar, i.e., it makes no difference in the quality of the algorithm if the data were recorded at a subject's wrist or hip.

The green dots in Figure 3.14 mark the amount of correct sleep and wake scorings at the differentiating threshold of an optimal calculation of total sleep time (TST). The TST measures the total time of NREM sleep and REM sleep in a nocturnal recording. For determining the

optimal threshold TST is computed for every possible value of the threshold for each of the 30 subjects. Ensuing, one chooses the specific threshold of every subject, whose corresponding TST is equal to the TST determined by polysomnography. The global optimal threshold for calculating TST is then obtained by averaging these 30 subject-specific thresholds. Table 3.2 shows the resulting thresholds and the corresponding percentages of correct sleep and awake scorings (shaded in green, identical with coordinates of the green dots in Figure 3.14) as well as the scorings of the manufacturer software (SOW). In addition to facilitate a comparison between the results for SOW and SAD, the coordinates of the ROCs from Figure 3.14 are given with either the sleep value or the awake value chosen identical to the one from SOW (see lines 2, 3, 5, 6).

In relation to SOW the SAD algorithm yields clearly better results. This is apparent by reference to the blue and red marked cells in a row which show the percentages of correct sleep or awake scorings obtained by SOW and SAD at a similar opposite percentage, i.e., if one compares percentages of correct sleep scorings the percentage of correct awake scoring is fixed (corresponding to the SOW result).

Tab. 3.2: Results of the receiver operating characteristic, i.e., the optimal threshold and the corresponding percentage of correct sleep and awake scoring (marked green) as well as a comparison of scorings obtained by the manufacturer software (SOW) and the SAD algorithm (blue and red cells in a row).

algorithm	wrist			hip		
	sleep	awake	threshold	sleep	awake	threshold
SOW	83.2	48.1	N/A	97.8	10.0	N/A
SAD _{NAT}	83.2	66.7	-6.3128	97.8	26.9	-5.4090
	91.3	48.1	-5,8402	99.9	10.0	-4.7622
opt. SAD _{NAT}	87.0	58.8	-6.1158	86.1	57.9	-6.0480
SAD _{NORM}	83.2	75.3	0.2862	97.8	37.9	0.7181
	95.8	48.1	0.8015	100.0	10.0	1.4328
opt. SAD _{NORM}	92.6	58.5	0.5978	92.3	58.9	0.4247

The green marked thresholds in Table 3.2 are used to calculate two sleep parameters, the total sleep time (TST) and the counts of awake during night after falling asleep (CAW). Each value of the classification function (corresponding to a 30-second accelerometric measurement) greater than the obtained threshold means a 30-second wake phase and all the other values imply a 30-second sleep state. Therefore the TST and the CAW are obtained by counting the respective 30-second intervals. The ascertained sleep parameters of every subject are compared with the polysomnographic results (TST: 375.3 ± 69.1 min; CAW: 23.1 ± 11.4) for evaluating the quality of the algorithm.

Tab. 3.3: Comparison of total sleep time (TST) and counts of awake (CAW) determined by the SAD algorithm, manufacturer software (SOW) and the subject's self-reporting using average deviation ($\langle|\Delta\text{TST}|\rangle$ & $\langle|\Delta\text{CAW}|\rangle$), standard deviation of parameter difference ($\sigma(|\Delta\text{TST}|)$ & $\sigma(|\Delta\text{CAW}|)$) and Pearson correlation coefficient r in comparison to polysomnography.

method	$\langle \Delta\text{TST} \rangle$	$\sigma(\Delta\text{TST})$	$r_{\text{PSG,meth}}$	$\langle \Delta\text{CAW} \rangle$	$\sigma(\Delta\text{CAW})$	$r_{\text{PSG,meth}}$
hip _{NAT}	49.5 min	39.1 min	0.388	6.6	5.7	0.642
wrist _{NAT}	46.2 min	37.2 min	0.533	7.8	7.6	0.274
hip _{NORM}	48.4 min	40.9 min	0.416	7.3	5.6	0.614
wrist _{NORM}	47.3 min	44.2 min	0.426	7.5	6.5	0.514
hip _{SOW}	83.1 min	75.9 min	0.268	21.8	11.1	0.395
wrist _{SOW}	70.5 min	74.7 min	0.396	16.4	11.1	0.297
self-reporting	82.2 min	75.1 min	0.585	N/A	N/A	N/A

Table 3.3 reports three parameters which are used for estimating the results of the SAD

algorithm, the manufacturer software (SOW) and the subject's self-reporting. The average deviation $\langle |\Delta x| \rangle$ and the standard deviation $\sigma^2(|\Delta x|)$ are calculated as follows:

$$\langle |\Delta x| \rangle = \frac{1}{N} \sum_{n=1}^N |x_n^{\text{PSG}} - x_n^{\text{method}}| \quad \sigma^2(\langle |\Delta x| \rangle) = \frac{1}{N-1} \sum_{n=1}^N [|x_n^{\text{PSG}} - x_n^{\text{method}}| - \langle |\Delta x| \rangle]^2$$

where $x \dots$ sleep parameter (TST or CAW)
 $N \dots$ number of subjects (here $N = 30$)

The Pearson correlation coefficient r is an indicator of the linear dependences between two time series; here it is used to study linear dependence between PSG and the other methods, SAD, SOW, and self-reporting) and is determined as described in Section 2.4.2.

All results shown in Table 3.3 are obtained after calibrating the sleep parameters TST and CAW. For this purpose the sleep parameters determined by polysomnography and by the SAD algorithm are compared in a two-dimensional plot (see Figure 3.15 for TST). Since the calculated values are fluctuating around the ideal correlation (black dashed line in Figure 3.15) the systematic error is removed by using a linear fit (see Figure 3.15, coloured lines and boxes). Due to this calibration the average deviation and the standard deviation are clearly reduced. The normalisation of slope, intercept and respiratory peak-ratio before determining the classification function yields a similar effect. Therefore, a further linear calibration of the normalised data yields no significant improvements.

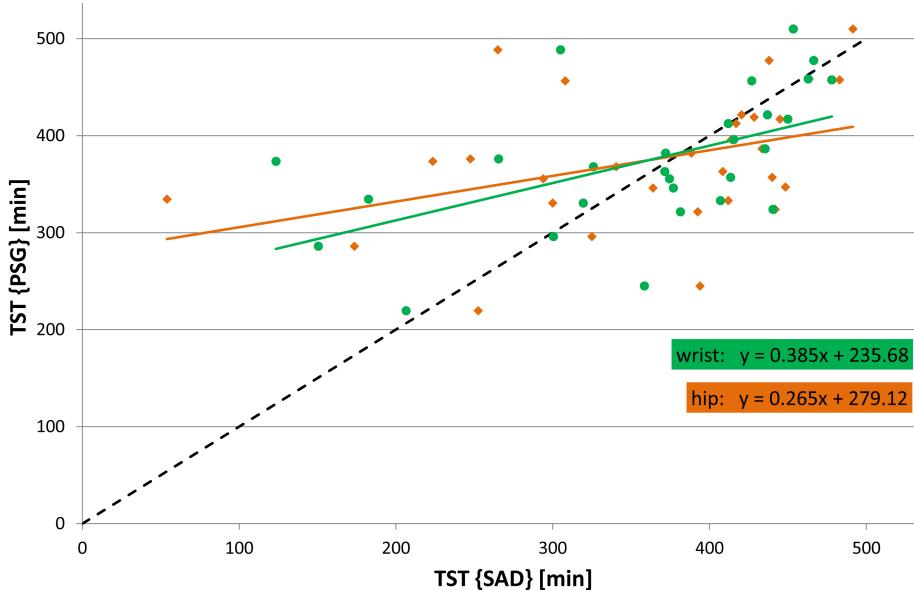


Fig. 3.15: TST calculated by polysomnography versus TST derived by the SAD algorithm (natural wrist data in green and natural hip data in orange). The linear fits for both data types in the same colours are used for removing the bias of the SAD algorithm. The black dashed line shows the identical relation.

It is apparent from Table 3.3 that the sleep parameters are more precisely determined by the SAD algorithm than by SOW or self-reporting. With the SAD algorithm the average deviation of TST is reduced by 25 minutes and the standard deviation of TST is almost divided in half compared with the SOW and the subject's self-reporting. A similar behaviour is observed for CAW, i.e., the average deviation is nearly divided by three and the standard deviation is divided in half. These results show that the SOW does not yield better results for calculating TST than the subject's self-reporting and it performs worse than the SAD algorithm. According to this analysis one also expects that a wrist measurement results in a more slightly precise TST and a measurement at a human hip yields slightly better results for CAW.

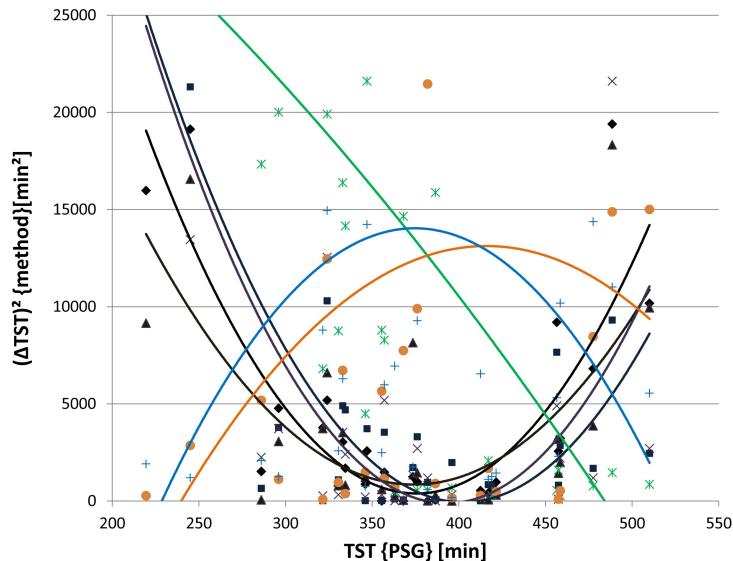


Fig. 3.16: Squared deviation $(\Delta TST)^2$ as a function of TST determined by polysomnography (dark curves - SAD, green (hip) and orange (wrist) curves - SOW, blue curve - self-reporting)

exhibits an opposite curvature than the other curves. The SAD algorithm (dark curves) has its optimum at average total sleep times and increases clearly at extreme values. However, the subject's self-reporting (blue curve) yields the best results (low variance) at very short and long total sleep times. That means, humans can estimate their TST very well if they sleep either briefly or for a longer than average time. The SOW (orange curve - wrist data) reveals a quite similar behaviour like the subject's self-reporting.

The behaviour of the CAW is shown in Figure 3.17. Again, the dark curves (results of the SAD algorithm) exhibit their optimum at average CAW and increase clearly at extreme values. Due to the green and orange curves one can conclude that the CAW is only calculated correctly by SOW if a subject is barely awake. I.e., if a human is often awake because of a sleep disorder or due to other diseases the manufacturer software does not yield correct values (compared to PSG). The results might suggest that the CAW calculation of SOW does not perform correctly because the SOW obtains no values larger than 7 and one half of all values is equal to zero (based on hip measurements in 30 subjects).

The sleep-aware differentiation based on spectral analysis of accelerometric recordings measured by SOMNOwatch can be used to construct a hypnogram similar to the traditional one from polysomnography. Figure 3.18 is a compact illustration of the hypnogram determined by polysomnography, the sleep-aware differentiation employing the four SAD algorithms,

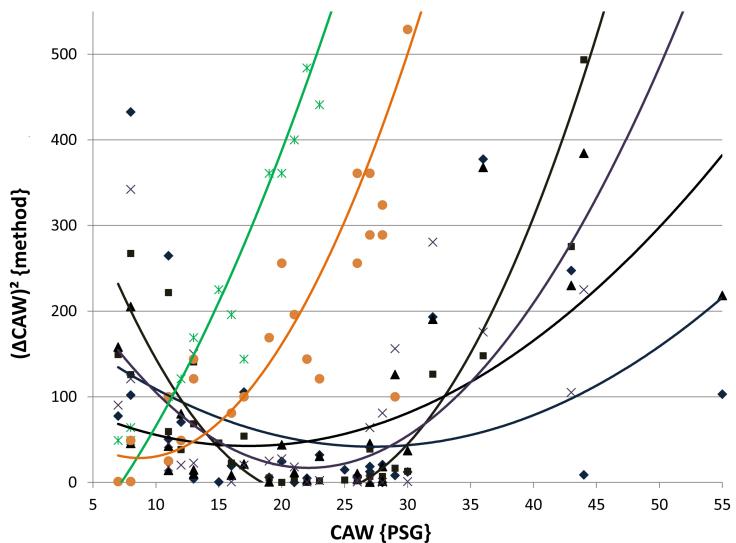


Fig. 3.17: Squared deviation $(\Delta CAW)^2$ as a function of CAW determined by polysomnography (dark curves - SAD, green (hip) and orange (wrist) curves - SOW)

In order to assess the behaviour of the sleep parameter deviation for each subject, the squared deviation is plotted versus the sleep parameter determined by polysomnography. Figure 3.16 shows the performance of the TST deviation calculated by the SAD algorithm (dark), SOW (green and orange), and self-reporting (blue). For facilitating the interpretation the results of each method are fitted by a second order polynomial function. Note that a linear fit is clearly insufficient in most cases here while data scatter too much for reliable fits with higher order polynomials.

It is obvious that the dark curves The squared TST deviation of the SAD algorithm (dark curves) has its optimum at average total sleep times and increases clearly at extreme values. However, the subject's self-reporting (blue curve) yields the best results (low variance) at very short and long total sleep times. That means, humans can estimate their TST very well if they sleep either briefly or for a longer than average time. The SOW (orange curve - wrist data) reveals a quite similar behaviour like the subject's self-reporting.

and the two frequency spectra based on accelerometric recordings from a subject's hip and wrist. One can see that the results of the SAD algorithms are in good accordance with the polysomnographic sleep classification. Furthermore, it is obvious that the SAD algorithms overestimate the length of wake phases due to the used smoothing procedures. This limitation should be further investigated and eliminated by improving the here introduced SAD algorithm.

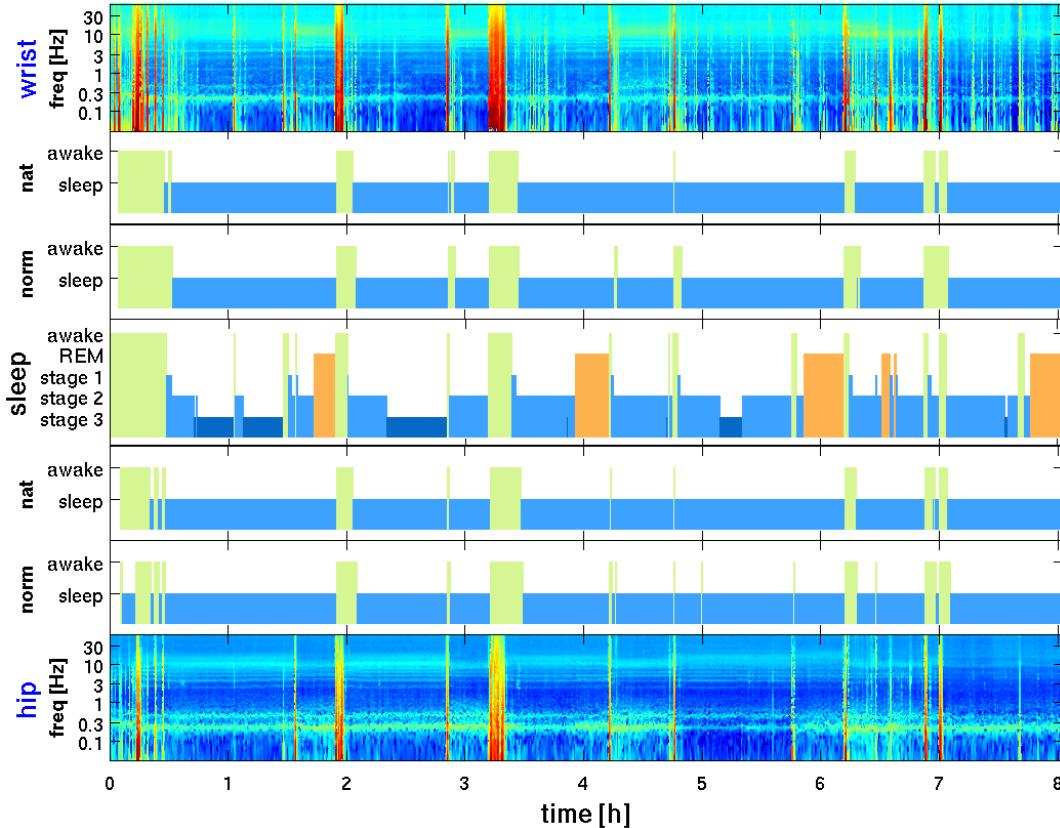


Fig. 3.18: Sleep-aware differentiation from nocturnal accelerometric recordings in a healthy subject. The upper three panels show the spectrogram and the SAD-based hypnograms for wrist raw data and wrist normalised data in comparison with the polysomnographic hypnogram in the center panel and the corresponding results obtained from hip recordings in the lower three panels.

3.3.4 Algorithm Test

The SAD algorithm for the SOMNOWatch device was developed by analysing nocturnal accelerometric recordings from 30 subjects of the P6 study. The measurements obtained from the remaining 70 subjects of the P6 study serve as the control group for this algorithm. The obtained SAD algorithm and the corresponding thresholds for differentiating sleep and wake phases in the training sample are now applied to these 70 recordings without further optimisation. With this independent sample the quality of the SAD algorithm can be checked. Figure 3.19 shows the ROC curves of wrist and hip measurements based on the training (30 subjects) and control group (70 subjects). These two samples yield quite similar results for the natural method of the SAD algorithm (see Figure 3.19, blue curves) thereby supporting the previous claims concerning the algorithm's capabilities. However, the normalised variant of the SAD algorithm (red curves) applied to the control sample shows a slightly worse performance compared with the training sample. The AUC of the wrist and hip based ROC curves are reduced by 3 and 5 percentage points. All ROC curves of the control sample exhibit a nearly similar AUC (wrist: 0.84 - 0.85, hip: 0.81 - 0.82).

Moreover, the green dots in Figure 3.19 show that the fixed thresholds correspond to quite different correct sleep and awake scorings contrary to the results described in Section 3.3.3.

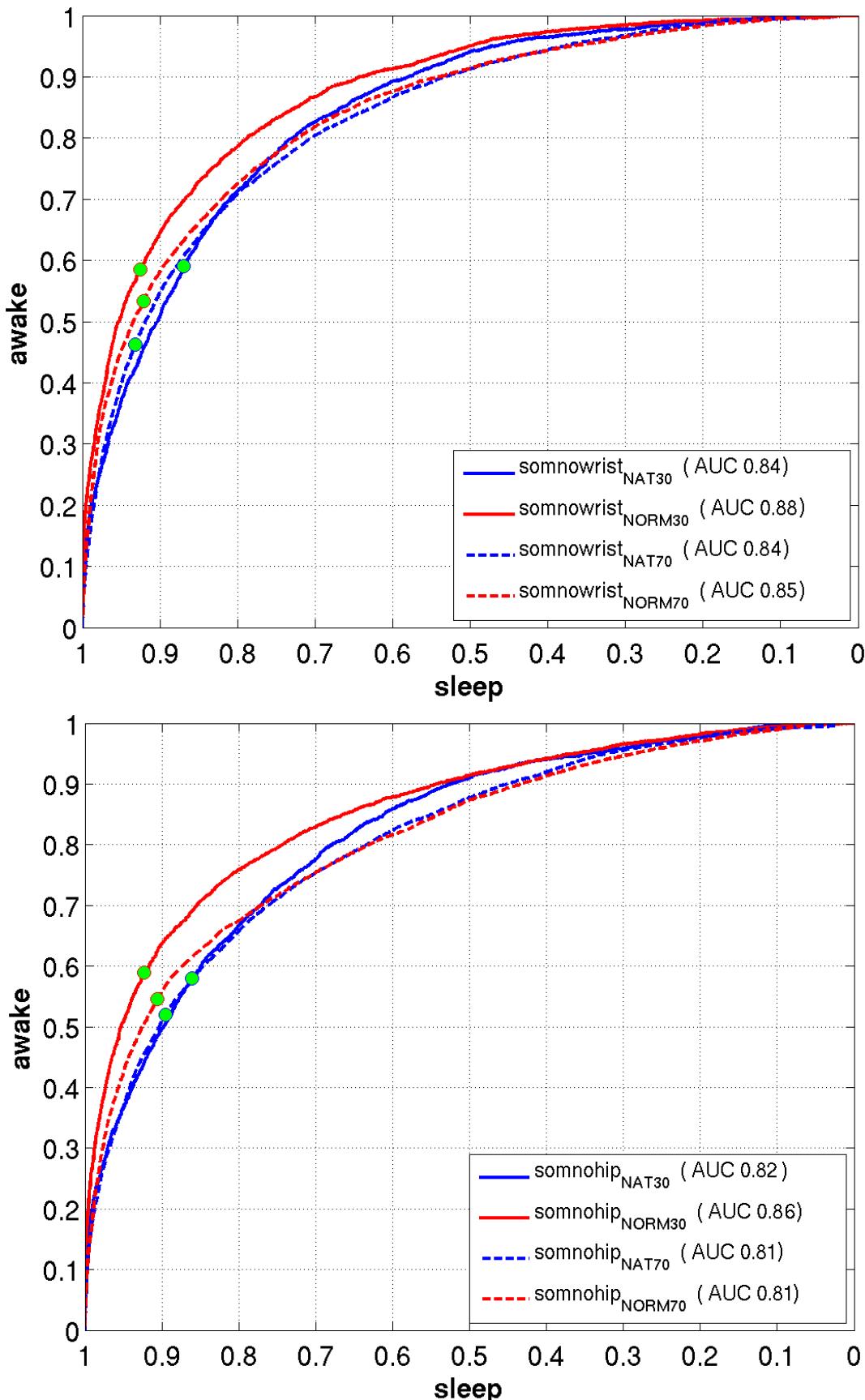


Fig. 3.19: Receiver operating characteristic of SAD algorithm for wrist (upper plot) and hip measurements (lower plot) of 30 and 70 subjects, where the correct awake scoring versus the correct sleep scoring (compared with the polysomnographic sleep classification) is illustrated. The final AUC values are reported in the legend. The green dots mark the amount of correct sleep and awake scorings at the optimal threshold for calculating the total sleep time.

Table 3.4 shows the different method's percentage reliability obtained from ROC curves and confirms the discussed conclusions.

Tab. 3.4: Same as Table 3.2 but for the control group of 70 subjects. Results of the receiver operating characteristic, i.e., the optimal threshold and the corresponding percentage of correct sleep and awake scoring (marked green) as well as a comparison of scorings obtained by the manufacturer software (SOW) and the SAD algorithm (blue and red cells in a row).

algorithm	wrist			hip		
	sleep	awake	threshold	sleep	awake	threshold
SOW	83.2	48.1	N/A	97.8	10.0	N/A
SAD _{NAT}	83.2	66.9	-6.5934	97.8	25.9	-5.5460
	92.6	48.1	-6.1616	99.8	10.0	-4.8436
opt. SAD _{NAT}	93.2	46.2	-6.1158	89.5	52.0	-6.0480
SAD _{NORM}	83.2	68.6	0.2911	97.8	29.5	0.8499
	94.1	48.1	0.7009	99.7	10.0	1.4242
opt. SAD _{NORM}	92.2	53.3	0.5978	90.6	54.6	0.4247

The TST calculation in the control group yields similar but somewhat better results when compared with a TST calculation in the training sample (see Table 3.3 and 3.5). Especially the SOW results have clearly improved. This can possibly be explained by the fact that the control group is more homogeneous than the training sample, i.e., there is a smaller number of subjects with an unusual sleep behaviour in the control group. The CAW computation using the SAD algorithm yields a slightly inferior result, but it is still better than the obtained CAW using SOW. This is an indication of an incorrect CAW calculation when using SOW.

Tab. 3.5: Same as Table 3.3 but for the control group of 70 subjects. Comparison of total sleep time (TST) and counts of awake (CAW) determined by the SAD algorithm, manufacturer software (SOW) and the subject's self-reporting using average deviation ($\langle|\Delta\text{TST}|\rangle$ & $\langle|\Delta\text{CAW}|\rangle$), standard deviation of parameter difference ($\sigma(|\Delta\text{TST}|)$ & $\sigma(|\Delta\text{CAW}|)$) and Pearson correlation coefficient r in comparison to polysomnography.

method	$\langle \Delta\text{TST} \rangle$	$\sigma(\Delta\text{TST})$	$r_{\text{PSG,meth}}$	$\langle \Delta\text{CAW} \rangle$	$\sigma(\Delta\text{CAW})$	$r_{\text{PSG,meth}}$
hip _{NAT}	43.6 min	41.6 min	0.539	8.3	8.5	0.269
wrist _{NAT}	41.4 min	39.4 min	0.676	8.1	8.4	0.296
hip _{NORM}	44.4 min	32.5 min	0.567	9.1	8.9	0.407
wrist _{NORM}	40.9 min	33.6 min	0.638	9.9	10.5	0.291
hip _{SOW}	81.6 min	44.9 min	0.708	22.9	11.3	0.473
wrist _{SOW}	57.3 min	60.6 min	0.429	18.5	11.3	0.355

Tab. 3.6: Mean value and standard deviation of Total sleep time (TST) and counts of awake (CAW) in the training sample and the control sample derived by polysomnography.

sample	$\langle\text{TST}\rangle$ [min]	$\sigma(\text{TST})$ [min]	$\langle\text{CAW}\rangle$	$\sigma(\text{CAW})$
training	375.3	69.1	23.1	11.4
control	361.0	66.4	24.1	12.0

4 Summary and Future Prospect

The aim of this research work was to identify human movement patterns from acceleration data measured within the framework of the German National Cohort pre-study I by employing three different accelerometric devices. To study the characteristics of the data and to assess signal quality, spectral analysis was applied.

The spectral analysis of nocturnal accelerometric recordings measured from 100 healthy subjects by SOMNOwatch has revealed several spectral features and led to the finding of a typical whole-night frequency spectrum. The expectation of stronger movements without a preferred frequency during wake phases compared with NREM sleep and REM sleep was confirmed. Furthermore, it was apparent that almost every transition from deep sleep to light sleep or wake is associated with a strong body movement, including arm movements or turns in bed. This analysis has also shown that traces of two physiological phenomena with a periodic behaviour, the human respiration (frequency range: 0.2 Hz - 0.4 Hz) and movements possibly associated with lung sounds caused by lung rattling or snoring (frequency range: 4 Hz - 10 Hz) can be found in accelerometric data. Further research shall address whether the latter physiological phenomenon is truly related to lung sounds or is the consequence of another body function.

When studying wake states and sleep stages in a double logarithmic power spectrum of nocturnal accelerometric recordings measured by SOMNOwatch notable differences between both states have been determined. The power-law behaviour which has been found for each curve differs in spectral exponent, i.e., the awake spectrum exhibits a f^{-2} behaviour and the spectra of the different sleep stages are proportional to f^{-1} . Additionally, the power-law trend during sleep is interrupted by two distinct peaks, the narrow respiratory peak followed by its higher harmonics and the indistinct wide peak in the frequency range from 4 Hz to 10 Hz. Significant differences between NREM sleep and REM sleep could not be determined in accelerometric data.

The comparison between day and night measurements has identified a very high amount of strong movements without a preferred frequency during daytime in contrast to bedtime, and it has confirmed the prevailing expectation of stronger movements during wake phases compared with NREM sleep and REM sleep. Additionally, it has been shown that time periods where the subject took off the accelerometer are characterised by a drastically reduced motion activity and missing respiratory signal. The implementation of an automated filtering algorithm will be a part of future research.

In the context of the German National Cohort pre-study I three different accelerometric devices - SOMNOwatch, Actigraph GT3Xplus and GENEactiv - are used for long-term (up to seven days) acceleration recordings. When comparing these three accelerometers the SOMNOwatch device came out as the best suited device for achieving the aims of the project. It has a superior signal-to-noise level and it does not introduce artifacts in the data due to power saving modes. In contrast, the GENEactiv device is affected by a high noise level (white noise) so that all spectral information above approximately 0.5 Hz are lost. Thus a sleep-aware differentiation based on spectral analysis is not possible. Due to a "feature" called "low power mode" (LPM) implemented by the Actigraph company, the Actigraph GT3Xplus accelerometer could not be used for developing a SAD algorithm, too. The reduction in sampling from 80 Hz to 1 Hz during approximately 94 percent of nighttime and the general structure of the LPM have hindered any data evaluation, analysis, and related derivations. After developing a firmware update the LPM's structure was modified and the time of suspension into LPM was clearly reduced (23 percent of the total recording time). Now, the revised Actigraph GT3Xplus can be used for the development of a spectral analysis based SAD algorithm. However, if the Actigraph GT3Xplus device is used in the main study

of the National Cohort, a further firmware update will be needed, because the data losses due to LPM/LPM2 are still not acceptable.

The basis of the developed SAD algorithm is the double logarithmic power spectrum. The classification function complies a linear combination of the three distinctive features slope, intercept, and respiratory peak-ratio optimised by two smoothing procedures (long-range backward EMA and short-range peaked EMA). For differentiating sleep and awake all values of the linear combination larger than a specific threshold mean the human is awake and all values lower than a specific threshold mean the human sleeps. This algorithm obtains clearly better results compared with the manufacturer software (SOW) and the subject's self-reporting which has been shown by calculating the total sleep time and the counts of wake states during night. These results were confirmed by applying the developed SAD algorithm to a larger control group of 70 subjects.

In conclusion, one can affirm that this work makes an important contribution to the main study of the German National Cohort starting in 2013.

5 List of References

List of Literature

- [1] COLE R.J.; KRIPKE D.F.; GRUEN W.; MULLANEY D.J.; GILLIN J.C.
Automatic sleep/wake identification from wrist activity.
 Sleep 15 (1992) 461-469
- [2] SADEH A.; SHARKEY K.M.; CARSKADON M.A.
Activity-based sleep-wake identification: an empirical test of methodological issues.
 Sleep 17 (1994) 201-207
- [3] GODFREY A.; CONWAY R.; MEAGHER D.; ÓLAIGHIN G.
Direct measurement of human movement by accelerometry.
 In: Medical Engineering & Physics 30 (2008) 1364-1386
- [4] MATHIE M.J.; COSTER A.C.F.; LOVELL N.H.; CELLER B.G.
Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement.
 In: Physiological Measurement 25 (2004) R1-R20
- [5] HASLEY W.D.
Macmillan dictionary for students.
 Macmillan Pan Ltd. (1984)
- [6] SILBER M.H.; ANCOLI-ISRAEL S.; BONNET M.H.; CHOKROVERTY S.; GRIGG-DAMBERGER M.M.; HIRSHKOWITZ M.; KAPEN S.; KRYGER M.H.; PENZEL T.; PRESSMAN M.R.; IBER C.
The visual scoring of sleep in adults.
 In: Journal of Clinical Sleep Medicine 3 (2007) 121-131
- [7] ERLACHER D.
Anleitung zum Klarträumen: Die nächtliche Traumwelt selbst gestalten.
 Books on Demand (2010) 9-13
- [8] NATIONAL COHORT
A prospective epidemiologic study resource for health and disease research in Germany.
 Project proposal dated from February 2011
- [9] WIENTZEK A.
Feasibility study - evaluation of the feasibility of mono- and multifunctional devices to objectively measure physical activity - National Cohort.
 Presentation at the 2nd GSISH Biosensor Workshop Munich 2011
- [10] METZ C.E.
Receiver Operating Characteristic analysis: A tool for the quantitative evaluation of observer performance and imaging systems.
 Journal of the American College of Radiology 3 (2006) 413-422
- [11] OBUCHOWSKI N.A.
Receiver Operating Characteristic curves and their use in radiology.
 Radiology 229 (2003) 3-8

- [12] BRONSTEIN I.N.; SEMENDJAJEW K.A.; MUSIOL G.; MÜHLIG H.
Taschenbuch der Mathematik.
Harri Deutsch Verlag (2008) 844-846
- [13] BUNDE A.; HAVLIN S.; KANTELHARDT J.W.; PENZEL T.; PETER J.-H.; VOIGT K.;
Correlated and uncorrelated regions in heart-rate fluctuations during sleep.
IEEE Transactions on Biomedical Engineering 29 (1982) 456-457
- [14] KOBAYASHI M.; MUSHA R.
1/f fluctuation of heartbeat period.
Physical Review Letters 85 (2000) 3736-3739
- [15] TAKAHASHI K.; KOYAMA Y.; KAYAMA M.W.
Is state-dependent alternation of slow dynamics in central single neurons during sleep present in the rat ventroposterior thalamic nucleus?
Neuroscience Research 48 (2004) 203-210
- [16] GILDEN D.L.; THORNTON T.; MALLON Y.; NAKAMURA K.; YAMAMOTO M.
1/f noise in human cognition
Science 267 (1995) 1837-1839
- [17] HAUSDORFF J.M.; PENG C.-K.; LADIN Z.; WEI J.Y.; GOLDBERGER A.L.
Is walking a random walk? Evidence for long-range correlations in stride interval of human gait
Journal of Applied Physiology 78 (1995) 349-358
- [18] GRAF C.E.; GIANNELLI S.V.; HERRMANN F.R.; SARASIN F.P.; MICHEL J.P.; ZEKRY D.; CHEVALLEY T.
Identification of older patients at risk of unplanned readmission after discharge from the emergency department - comparison of two screening tools.
Swiss Medical Weekly (2012) doi: 10.4414/smw.2012.13327

List of Figures

2.1	Schematic diagram of a spring/mass system	7
2.2	Vectorial composition of measured signal	7
2.3	Measured signal of a triaxial hip-mounted accelerometer during sleep	8
2.4	Schematic hypnogram of a whole night of a healthy subject with four sleep cycles	9
2.5	Three main accelerometers (www.somnomedics.eu , www.theactigraph.com , www.geneactiv.co.uk - extracted on 12.06.2012)	11
2.6	Function of time $f(t) = \vartheta(t+3) - \vartheta(t-3)$	12
2.7	Function of frequency $g(\omega) = \sqrt{\frac{2}{\pi}} \frac{\sin(3\omega)}{\omega}$	12
2.8	Unbinned double logarithmic power spectrum	13
2.9	Binned double logarithmic power spectrum	13
2.10	Diagnostic indices for SAD	15
2.11	ROC of a SAD with accelerometry	15
3.1	Frequency spectrum of acceleration recordings of a whole night	17
3.2	Calibration curve for colour coding of power spectra. For different spectral intensities (horizontal axis) and different frequencies (vertical axis) a logarithmic colour coding is chosen from low intensity (blue) to high intensity (red)	18
3.3	Power spectrum of a nocturnal accelerometric measurement	19
3.4	24h spectrogram calculated from an accelerometric recording	21
3.5	Accelerations at a subject's hip measured by three different accelerometers during a 10-second episode of quiet sleep with very weak activity	22
3.6	Power spectra of accelerometric recordings measured by SOMNOwatch and GENEactiv in six different subjects	23
3.7	Accelerations at a subject's hip measured by SOMNOwatch (left) and Actigraph GT3Xplus (right) during the same motion event (data from P6 study).	24
3.8	Accelerations at a subject's hip measured by Actigraph GT3Xplus	25
3.9	Semi-logarithmic power spectrum of an exemplary nocturnal accelerometric recording measured by Actigraph GT3Xplus	25
3.10	Accelerations at a subject's hip measured by the revised Actigraph device during sleep	26
3.11	Power spectra of accelerometric recordings measured by SOMNOwatch, GENEactiv and Actigraph GT3Xplus (after firmware update) in six different subjects	27
3.12	Power spectrum as basic of a SAD algorithm	28
3.13	Frequency spectrum and parameter behaviour of a nocturnal acceleration recording measured by SOMNOwatch	30
3.14	Receiver operating characteristic of four different SAD algorithms	32
3.15	Calibration curve of TST	34
3.16	Squared deviation $(\Delta TST)^2$ as a function of TST determined by polysomnography	35
3.17	Squared deviation $(\Delta CAW)^2$ as a function of CAW determined by polysomnography	35
3.18	Sleep-aware differentiation from nocturnal accelerometric measurements in a healthy subject	36
3.19	Receiver operating characteristic of SAD algorithm for wrist and hip measurements of 30 and 70 subjects	37

List of Tables

2.1	Main features of sleep and wake states	9
2.2	The main features of the accelerometers Actigraph GT3Xplus, GENActiv, and SOMNOWatch	11
2.3	Two different methods of exponential moving average (EMA)	14
3.1	Statistical analysis of six 7-day Actigraph GT3Xplus recordings from P4 . .	24
3.2	Results of the receiver operating characteristic (30 subjects)	33
3.3	Comparison of total sleep time (TST) and counts of awake (CAW) determined by the SAD algorithm, manufacturer software (SOW) and the subject's self-reporting	33
3.4	Results of the receiver operating characteristic (70 subjects)	38
3.5	Comparison of total sleep time (TST) and counts of awake (CAW) calculated by the SAD algorithm and SOW (70 subjects)	38
3.6	Mean value and standard deviation of Total sleep time (TST) and counts of awake (CAW) in the training sample and the control sample derived by polysomnography	38

6 Statement of Authorship

I, Patrick Wohlfahrt, hereby take an affidavit that the work presented in this bachelor thesis has been performed solely by myself and only by using the stated resources. All cogitations and results obtained from literature are indicated.

I confirm that this work is submitted in partial fulfilment for the degree of Bachelor of Science (BSc) in Medical Physics and has not been submitted elsewhere in any other form for the fulfilment of any other degree or qualification.

Date, place

Patrick Wohlfahrt

7 Acknowledgement

I would like to thank my advisor PD. Dr. rer. nat. Jan W. Kantelhardt (institute of physics, MLU Halle) for his expert advices on the theoretical and practical aspects of my work, for his unending enthusiasm and persistent motivation as well as for all hours of trustful cooperation. His outstanding professional competence and excellent teaching expanded my knowledge and improved my personal skills.

Furthermore I would like to acknowledge Melanie Zinkhan and Prof. Dr. med. Andreas Stang (institute of clinical epidemiology, MLU Halle), who prepared the complete data sets I used in my work, answered my questions and contributed useful ideas and approaches in many discussions towards the fulfillment of this project.

A special thanks to Dr. Aicko Yves Schumann and Dr. Claire Christensen (Complexity Science Group, University of Calgary) for many hours of proofreading my bachelor thesis and numerous helpful hints for improving my scientific writing in general and especially in English.

I would also like to thank PhD Mirko Kämpf (institute of physics, MLU Halle) and my two fellow students Berit Schreck and Arne Böker for many methodical and fruitful discussions, for their assistance as well as for the very joyful working atmosphere in our office.

Finally, many thanks to my family, friends and all the other people who supported me during this research.