Dissertation

Novel sleep screening based on tracheal body sound and actigraphy

Dissertation to obtain the doctoral degree of

Human Biology (Dr. biol. hum.)

at the

Medical Faculty of Ulm University

by

Christoph Kalkbrenner

from Leipzig

2018
Acting dean: Prof. Dr. rer. nat. Thomas Wirth
1. Reviewer: Prof. Dr. med. Wolfgang Rottbauer
2. Reviewer: Prof. Dr. med Tibor Kesztyüs

Date of graduation: 29th June, 2018
Contents

Abbreviations vii

1 Introduction 1
  1.1 Problem description . . . . . . . . . . . . . . . . . . . . 1
  1.2 Medical background . . . . . . . . . . . . . . . . . . . . 3
    1.2.1 Obstructive sleep apnea . . . . . . . . . . . . . . 3
    1.2.2 Sleep stages . . . . . . . . . . . . . . . . . . . . 5
  1.3 Technical background . . . . . . . . . . . . . . . . . . 7
    1.3.1 Polysomnography . . . . . . . . . . . . . . . . . . 7
    1.3.2 Bioacoustics . . . . . . . . . . . . . . . . . . . . 9
    1.3.3 Ambulant sleep monitoring . . . . . . . . . . . . . 10
    1.3.4 Related work . . . . . . . . . . . . . . . . . . . . 11
  1.4 Objective of this work . . . . . . . . . . . . . . . . . . 13

2 Materials and Methods 15
  2.1 Hardware development . . . . . . . . . . . . . . . . . . 15
Contents

2.1.1 Recording body sound .................................. 17
2.1.2 Monitor setup ........................................ 19
2.1.3 Data interface ........................................ 20
2.2 Data Processing ....................................... 21
  2.2.1 Apnea–hypopnea index calculation ................. 21
  2.2.2 Heart rate calculation ............................. 28
  2.2.3 Movements and sleeping position ................. 31
  2.2.4 Sleep staging ..................................... 32
2.3 Clinical trial ........................................... 36
  2.3.1 Clinical trial registration .......................... 36
  2.3.2 Study design ...................................... 37
  2.3.3 Subjects ........................................... 37
  2.3.4 Data acquisition .................................. 41
  2.3.5 Statistics .......................................... 42

3 Results .................................................. 45
  3.1 Apnea–hypopnea index ................................. 45
    3.1.1 Equivalence ..................................... 45
    3.1.2 Correlation and Bland-Altmann Analysis .......... 46
    3.1.3 Sensitivity and Specificity ...................... 48
  3.2 Sleep staging ........................................ 51
    3.2.1 Staging performance ............................ 51
    3.2.2 Sleep related parameters ....................... 52
Abbreviations

AHI  apnea-hypopnea index
ASM  ambulant sleep monitor
AUC  area under the curve
BMI  body mass index
CV   coefficient of variation
ECG  electrocardiography
EEG  electroencephalography
EMG  electromyography
EOG  electrooculography
EST  estimated
IIT  investigator initiated trial
IMU  inertial measurement unit
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDC</td>
<td>linear discriminant classifier</td>
</tr>
<tr>
<td>OSA</td>
<td>obstructive sleep apnea</td>
</tr>
<tr>
<td>PAS</td>
<td>possible apnea segment</td>
</tr>
<tr>
<td>PSG</td>
<td>in-laboratory polysomnography</td>
</tr>
<tr>
<td>RIP</td>
<td>respiratory inductance plethysmography</td>
</tr>
<tr>
<td>ROC</td>
<td>receiver operating characteristic</td>
</tr>
<tr>
<td>SD</td>
<td>standard deviation</td>
</tr>
<tr>
<td>SE</td>
<td>sleep efficiency</td>
</tr>
<tr>
<td>TST</td>
<td>total sleep time</td>
</tr>
<tr>
<td>TWT</td>
<td>total wake time</td>
</tr>
<tr>
<td>WASO</td>
<td>wake after sleep onset</td>
</tr>
</tbody>
</table>
Parts of this work have already been published in:


1 Introduction

The following chapters first describe the basic problem and also the motivation for this work. Subsequently, the essential medical and technical basics are explained. Based on this background information the objectives of this work can be derived and are stated.

1.1 Problem description

The American Academy of Sleep Medicine recognizes sleep disorders as a very common disease [5]. Moreover, the number of individuals suffering from sleep disorders is increasing at an alarming rate worldwide [13]. Regrettably, public recognition of the need for sufficient sleep, and awareness of sleep related conditions is poor [13]. The problem is intensified by the extensive process of a reliable sleep diagnosis. The gold standard for sleep diagnosis is the in-laboratory polysomnography (PSG) including a preceding sleep medicine interview and a physical examination.
1 Introduction

The labor, cost and time intensive nature of PSG, paired with the increasing prevalence of sleep disorders, has led to a strong demand for sleep laboratories. Additionally, numerous follow-up studies to track the effectiveness of an initiated therapy are also carried out using PSG [53]. This causes sleep centers worldwide to constantly operate at full capacity, which is reflected in long waiting lists typically exceeding six months. Moreover, patients are often reluctant to carry out a PSG, since the overnight stay in an unfamiliar sleep laboratory and the extensive recording equipment can have a significant impact on their already poor sleep. The high prevalence of sleep disorders and the enormous demand for investigations lead to the assumption that PSG cannot be used in all patients, which leads to an increasing number of untreated and undiagnosed sleep disorders.
1 Introduction

1.2 Medical background

1.2.1 Obstructive sleep apnea

A) open airways

B) obstructed airways

Figure 1: Nasal breathing: A) open airways, B) obstructed airways causing a typical breathing cessation patients with obstructive sleep apnea might suffer from. Derived from [14].

Sleep apnea is a highly prevalent respiratory sleep related disorder characterized by multiple breathing cessations during the night. There are three different forms of sleep apnea. obstructive sleep apnea (OSA) is an apnea in which the upper airways are closed during inhalation due to the negative pressure in the pharyngeal cavity and reduced innervation of the pharyngeal muscles as shown in Figure 1. Here, the breathing movements (thorax and abdomen) continue to be carried out further during the apnea. An apnea caused by obstruction is normally terminated with an arousal. In sleep medicine, an arousal represents the wake-up
reaction of the body. Arousals lead to an end of breathing cessations and thus prevent suffocation but reduce the recovery value of sleep. In contrast, a central apnea describes a loss of breathing due to impaired cerebral control of breathing. This also causes the breathing movements to fail. Mixed apnea describes the occurrence of both central apnea and obstructive apnea. With an estimated prevalence of 24 percent in males and 6 percent in females both aged 30 to 60 [74] and with 75 percent undiagnosed or untreated patients [75], OSA is the most common form of sleep apnea and considered a public health problem of the first order. OSA can lead to extensive daytime sleepiness [57], cognitive-behavioral problems and an elevated risk for cardiovascular disease [8, 12, 69] if left untreated. Even though current research questions the association of sleep apnea with cardiovascular events [17, 76], the resulting cognitive impairment still comes with personal and societal consequences, such as driving and workplace accidents, and a reduced quality of life.

The main criteria used to quantify the severity of OSA is the apnea-hypopnea index (AHI), which describes the mean number of breathing pauses longer than 10 seconds per hour during sleep. Breathing pauses are divided into the categories apnea and hypopnea. Apneas are defined by at least 90 % reduction in air flow and hypopneas are defined by at least 30 % reduction in air flow, including an event related arousal and/or more than 3 % oxygen desaturation [4]. The diagnosis can be confirmed for an AHI of 5 or greater, associated with respective symp-
1 Introduction

Symptoms not explained by other causes. Based on [3], the severity of OSA is defined as mild for 5-14 AHI, as moderate for 15-30 AHI, and as severe for an AHI greater than 30. For the measurement of AHI the PSG is the gold standard.

1.2.2 Sleep stages

As part of an objective sleep diagnosis, the monitoring of sleep stages during the night is of great importance. With this information the total sleep time (TST), the overall level of sleep efficiency (SE), and sleep disruptions can be evaluated. The distribution of sleep stages is required to

Figure 2: Hypnogram of an adult over the duration of one night manually created using in-laboratory polysomnography. REM, rapid-eye movement sleep stage; N1, N2, N3, non-REM sleep stages.
diagnose certain sleep disorders such as insomnia and circadian rhythm disorders [5]. According to the American Academy of Sleep Medicine, human sleep can be classified into the stages wake (W), rapid-eye movement (REM) and three non-REM stages (N1, N2, N3) [4]. N1 and N2 can be grouped into so called “light sleep” and N3 can be referred to as “deep sleep” [66]. For the assessment of sleep stages the overnight PSG is the gold standard. Here, electroencephalography (EEG), submental electromyography (EMG) and unilateral anterior tibial EMG, as well as electrooculography (EOG), are carried out in order to assess sleep stages. In a time consuming process the signals are manually evaluated by a trained technician who creates a so called hypnogram. Each 30 second time interval, a so-called epoch, is analyzed and assigned to one sleep stage. If two or more stages coexist during one epoch, the major sleep stage is scored. Figure 2 shows an example of a hypnogram created using PSG.
1 Introduction

1.3 Technical background

1.3.1 Polysomnography

Table 1: Extract of the most important recommended criteria for the evaluation of polysomnographic signals in the sleep laboratory according to [4, 51]. AHI, apnea–hypopnea index; EEG, electroencephalography; EOG, electrooculography; EMG, electromyography; ECG, electrocardiography.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Analyse</th>
</tr>
</thead>
<tbody>
<tr>
<td>airflow &amp; effort</td>
<td>• Number of obstructive/central/mixed apneas</td>
</tr>
<tr>
<td></td>
<td>• Number of hypopneas</td>
</tr>
<tr>
<td></td>
<td>• AHI</td>
</tr>
<tr>
<td></td>
<td>• Number of respiratory effort related arousals</td>
</tr>
<tr>
<td></td>
<td>• Breathing pattern (cheyne stokes, hypoventilation)</td>
</tr>
<tr>
<td>oxygen saturation</td>
<td>• Number of oxygen desaturations $&gt; 3 – 4%$</td>
</tr>
<tr>
<td></td>
<td>• Mean continuous oxygen saturation</td>
</tr>
<tr>
<td></td>
<td>• Minimum oxygen saturation during sleep</td>
</tr>
<tr>
<td>EEG, EOG, chin &amp; leg EMG</td>
<td>• Visual scoring of sleep stages</td>
</tr>
<tr>
<td></td>
<td>• Time in each stage</td>
</tr>
<tr>
<td></td>
<td>• Total sleep time (TST)</td>
</tr>
<tr>
<td></td>
<td>• Sleep latency</td>
</tr>
<tr>
<td></td>
<td>• Wake after sleep onset (WASO)</td>
</tr>
<tr>
<td></td>
<td>• Number of arousels</td>
</tr>
<tr>
<td></td>
<td>• Number of periodic limb/leg movements</td>
</tr>
<tr>
<td>ECG</td>
<td>• Average heart rate during sleep</td>
</tr>
<tr>
<td></td>
<td>• Highest heart rate during sleep</td>
</tr>
<tr>
<td>body position</td>
<td>• Number of apnoeas in each sleeping position</td>
</tr>
</tbody>
</table>

For the diagnosis of sleep disorders, the PSG in a sleep laboratory is the reference method. Under laboratory conditions, different biosignals
are recorded, stored, and evaluated over the period of one night. If the
disease to be diagnosed is a respiratory disorder, the internal pneumol-
ogy sleep laboratory is used. In order to diagnose sleep apnea, the
cardiorespiratory functions and the associated changes in neurophysi-
ological parameters (e.g. the occurrence of an arousal) are investigated.
The biosignals to be recorded here are based on the guideline of the
German Society for Sleep Research, or the American Academy of Sleep
Medicine, and are listed in Table 1. In addition to the listed biosignals,
a videometry of the patient is usually carried out. For the detection of
apnea, hypopnea, and arousal as well as for the determination of sleep
phases, the recorded data is manually evaluated by the medical staff the
next day.
1.3.2 Bioacoustics

Table 2: Relevant bioacoustic signals for sleep monitoring [23, 55].

<table>
<thead>
<tr>
<th>Group</th>
<th>Sub-group</th>
<th>Frequency band</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>lung sounds</td>
<td>trachea</td>
<td>0 – 2000</td>
<td>high volume</td>
</tr>
<tr>
<td></td>
<td>vesicular</td>
<td>0 – 600</td>
<td>audible at certain points on the chest</td>
</tr>
<tr>
<td></td>
<td>bronchial</td>
<td>0 – 600</td>
<td>audible near the central airways</td>
</tr>
<tr>
<td></td>
<td>broncho-vesicular</td>
<td>0 – 600</td>
<td>audible between lung and central airway</td>
</tr>
<tr>
<td>heart sounds</td>
<td>first heart sound</td>
<td>0 – 150</td>
<td>occurs at the beginning of the systole. Over 95% of signal energy below 75 Hz</td>
</tr>
<tr>
<td></td>
<td>second heart sound</td>
<td>0 – 150</td>
<td>occurs at the end of the systole. Over 99% of the signal energy below 75 Hz</td>
</tr>
<tr>
<td>snoring</td>
<td></td>
<td>30 – 250</td>
<td>mainly occurs during inspiration with a small expiratory component</td>
</tr>
</tbody>
</table>

The human body generates acoustic signals through the activity of the organs. This bioacoustic information is usually monitored in medicine for diagnostic purposes by means of a stethoscope (auscultation). Mostly heart sounds (e.g. heart valve sounds, heart tones), lung sounds (e.g. breathing sounds, lung noise), intestinal noises (e.g. intestinal activity) and blood vessel noises (e.g. flow sounds) are examined. Lung sounds might be a particularly valuable diagnostic parameter for the topic of sleep monitoring. It is assumed that after the successful recording of the lung
1 Introduction

sounds, the respiratory frequency, flow, apnea, and snoring can be identified. Furthermore, heart sounds are also interesting for sleep monitoring since heart rate variability is an indicator for sleep stages. Table 2 shows the most important bioacoustic signals for sleep monitoring.

1.3.3 Ambulant sleep monitoring

To address the difficulties described in Section 1.1, several ambulatory recording devices have been developed. Those devices mostly focus on the diagnosis of the most common cardiorespiratory sleep disorder, OSA. Here, more extensive ambulatory recording devices include four or more recording channels and are referred to as type 3 sleep monitors. These more complex devices often show strong performance in the diagnosis of OSA [19, 43, 49], but also come with some disadvantages. The convoluted setup is likely to cause data loss, especially if performed without medical attendance in a home setup. Additionally, the extensive recording equipment also has a significant impact on sleep quality. To overcome those limitations, extremely simplified ambulatory devices have been developed and are referred to as type 4 sleep monitors. These devices only include 2 or less recording channels and therefore benefit from a low price, simple setup, and can often be used in a home setting without medical assistance. These devices mostly utilize either nasal airflow or SpO2, or a combination of both [16, 53] to diagnose OSA. However,
these signals induce several problems and limitations. Essential diagnostic information about sleep time, the level of sleep quality and sleep disruptions cannot be measured. Due to mouth-breathing or misplacement, nasal airflow can frequently fail to measure breathing. Moreover, systematic reviews for the diagnostic performance of these monitors revealed poor results [64] and they are generally not recommended for a diagnostic test without additional PSG [53].

1.3.4 Related work

To overcome the limitations of recording nasal airflow (see Section 1.3.3) and to improve comfort and accessibility (see Section 1.1) for the patient, multiple systems have been proposed which instead use breathing sounds as main signal for measuring breath. Here, breathing sounds are recorded either by ambient air microphones located in the vicinity of the patient [6, 32], or by deploying special body sound microphones placed on the patient’s neck [47, 72]. These studies revealed a good performance in breath measuring. However, the effectiveness for sleep diagnosis is limited if breathing is the only diagnostic criteria. With the addition of more recording channels (ECG, EEG, pulse oximetry) this problem could be solved. However, these additions contradict sleep monitoring at a low price, and simple setup without medical assistance. It is suggested that body sound signals have a huge potential in sleep diagnosis but have not
1 Introduction

yet been fully exploited.

Additionally, several ambulant and automated approaches have also been developed, which do not measure breath but instead determine sleep stages and also avoid the disadvantages of PSG. To keep those devices simple and accessible, complex measuring methods like EEG, EOG and EMG are avoided. Instead, the variation in heart and breathing rate is measured in order to perform cardiorespiratory automated sleep stage classification. This topic has been extensively studied in recent years [59, 60, 71] and provides promising results. Here, an electrocardiography (ECG) is used to extract cardiac features, and standards like respiratory inductance plethysmography (RIP) are used to extract respiratory features. Furthermore, studies solely relying on respiratory features to assess sleep stages still showed acceptable correlation (> 70 %) with sleep stage classification compared to PSG [27, 39, 40].
1 Introduction

1.4 Objective of this work

This work aims to utilize and refine body sound technology to preserve simplicity while performing reliable sleep monitoring, including sleep staging for OSA diagnosis, a preliminary screening, or monitoring therapy effectiveness. It is suggested that body sounds can be used to detect breath sounds and heart sounds utilizing only a single microphone. This would allow a simple setup and high comfort, reducing the effect on sleep quality in comparison to PSG. Additionally, it is proposed that cardiorespiratory information extracted from the body sound signal in addition to actigraphy can be used to perform sleep staging. At the moment, sleep staging based purely on cardiorespiratory signals is only verified for well established recording methods (e.g. ECG) [41]. Furthermore, a fully automatic evaluation of the recorded signals is planned in order to save time and to allow diagnosis for non-sleep experts like family doctors. In order to test the developed hardware and software solutions, a study based on the gold standard PSG should be designed and conducted. If successful, this minimalistic approach could address the need for simple yet reliable sleep screening, including sleep staging and OSA diagnosis in an ambulatory setting. The following points summarize the aims of this work:

- Development of a sleep monitor based on body sound and actigraphy
- Development of an algorithm to automatically perform OSA diagno-
sis (based on body sound and actigraphy)

- Development of an algorithm to automatically perform sleep staging (based on body sound and actigraphy)
- Conduct a study to test previously developed hard- and software against the standard PSG
2 Materials and Methods

The following sections will first describe the hardware concept and development of the proposed sleep monitor. Based on this monitor two new and fully automated algorithms are described. These algorithms are designed to perform OSA diagnosis and sleep staging. Finally, to validate the proposed methods the conducted study based on the gold standard PSG is characterized.

2.1 Hardware development

To overcome the limitations mentioned in Section 1.1, a new sleep monitoring device has been developed which tries to preserve simplicity while performing reliable sleep monitoring. This device is referred to as an ambulant sleep monitor (ASM). As described in Chapter 1.3.2, measuring body sounds can offer information about breathing and heart activity in a very subtle way, and has therefore been chosen as the main method for
acquiring vital parameters during the night. In addition, an inertial measurement unit (IMU) is also included since it comes at a minimal cost, does not affect the patient’s sleep, and provides substantial information about the patient’s movements during the night. Figure 3 shows the hardware concept based on these sensors. The patient’s device is the actual ASM, a peripheral module which can be placed on the patient’s body. The data recorded with the ASM is then transferred via Bluetooth to the second device, for example a generic laptop or smartphone where the data is saved and processed. A more detailed look at the specifications of every component included in the developed hardware is given in the following sections.
2 Materials and Methods

2.1.1 Recording body sound

![Exploded view of the developed body sound microphone](image)

Figure 4: Exploded view of the developed body sound microphone

The most important part of the ASM design is the body sound recording. Commercially available electronic stethoscopes suit the task but are generally too expensive, too big and are not designed for overnight fixation. Body sound microphones used in published studies mostly consist of irreproducible and barely described prototypes like in [22, 47, 63] to name just a few examples. The only commercially available miniaturized body sound microphone is part of a system called LEOSound [35], which allows for long-term monitoring of lung sounds for the diagnosis of breathing disorders, like asthma. However, it also comes at a very high price, offers no interface for developers, and sleep monitoring is not included in
its field of application. Therefore, a well documented, reliable and comfortable method to record body sound has been specially developed.

Figure 4 illustrates this unique recording device offering optimized body sound extraction and noise suppression while maintaining a small size. For hygiene aspects, a disposable biocompatible single-use double-sided adhesive membrane is utilized to fix the microphone on the body for the entire night. To amplify the sound emitted by the human body, a cone shaped mechanical piece is placed in front of the microphone. For its high recording quality, compact geometry and low cost, an electret microphone is used to perform the actual body sound recording. To reduce external noise and movement artifacts, the microphone is placed within a flexible o-ring. For further signal enhancement, an appropriate analogue processing circuit is included at the front end. At first the signal is amplified to ensure that even shallow and therefore quiet breathing can be detected. Additionally, very low and very high frequency parts of the signal are removed to diminish disturbances from contact noise or aliasing effects during sampling. Therefore, signals in the frequency range from 30 to 2000 Hz are specially amplified and sampled. This is sufficient to capture the major parts of heart and breathing sounds referring to [50].
2.1.2 Monitor setup

Figure 5: Fixation of the microphone on patient’s neck in close vicinity to the trachea with the remaining hardware fixed on to a chest belt. IMU, inertial measurement unit.

Figure 5 illustrates the setup for the overnight screening. The developed ASM consists of a small case containing the miniaturized electronics, including IMU, battery, microcontroller, Bluetooth gateway and an inductive charging interface. Like most ambulatory sleep monitors the electronics are positioned on the patient’s chest using a flexible chest belt. A single lead connects the electronics with the specially designed body sound microphone. For acquiring both breathing and heart sounds, the selection...
2 Materials and Methods

of an optimal recording position for the body sound microphone is crucial. Several locations on the upper body and neck were considered. In agreement with [50] and previously published similar approaches [22, 47, 63], the optimal position for recording breathing sounds was found at the neck, in close vicinity to the trachea. In this position, muffled cardiac sounds can also be recorded, allowing the acquisition of heart beats.

2.1.3 Data interface

Data from both the microphone and IMU are sampled by the embedded system of the ASM. The analog audio signal is sampled with 5 kHz and 10 bits resolution, whereas the digital IMU signal is sampled with 250 Hz and 16 bit resolution. For the purpose of data reception and storage, a special software application was developed. The ASM sends data in realtime via Bluetooth to a laptop running said software. In order to ensure sufficient signal quality during initial manual positioning of the body sound microphone, the software displays the realtime audio data in its time and frequency domain. For data storage, the standardized data type called European Data Format (EDF) [33] was chosen. EDF facilitates the synchronous retention of multiple signals with varying sampling rates and is often referred to as standard for PSG recordings.
2.2 Data Processing

The following sections present the algorithms which were developed for automated AHI calculation, heart rate calculation and sleep staging, based on body sound and movement data. In this context the different signal processing methods are explained and examples are given to facilitate understanding of the algorithm sequences.

2.2.1 Apnea–hypopnea index calculation

In order to calculate AHI, all apneas and hypopneas over the course of a night need to be identified. The recorded body sound can be used to derive a representation of actual breath and flow, and thus drops in flow can be identified to detect apneas and hypopneas. To avoid a time-consuming manual evaluation, a fully automated algorithm has been developed to perform this task. Here, only a brief overview of the algorithm is given and a more detailed explanation can be found in [31]. The entire algorithm is split into the three basic segments called preprocessing, drop detection and classification. Figure 6 illustrates a simplified flow chart highlighting all major steps of the algorithm.
2 Materials and Methods

Figure 6: Flow chart including all major steps of the developed automated apnea/hypopnea detection. $E_1$: short-term breathing cycles, $E_2$: long-term amplitude changes, $E_3$: audio event detection, $PAS$: possible apnea segment, $e$: audio events, $AS$: apnea/hypopnea segment, $RS$: reference segment, $r_{hyp}$: reference value for hypopnea detection, $r_{ap}$: reference value for apnea detection, $AS_e$: audio events in $AS$, $RS_e$: audio events in $RS$. 

- Raw body sound
  - FIR bandpass 200 - 2000 Hz
  - Compute $E_1$ (breathing cycles)
  - Cut-off high amplitudes
  - Compute $E_2$ (long-term changes)
  - $E_2 <$ threshold
    - Yes: $E_1$ used
    - No: new PAS found
  - Compute $E_3$ (event detection)
  - Detect audio events $e$
  - Detect & remove $e$ caused by movements
  - Split PAS using $E_2$ in $AS$ & $RS$
  - Calculate reference $r_{hyp}$ & $r_{ap}$ using $e_{RS}$
  - Any period ($e_{AS} < r_{hyp}$) > 10s?
    - Yes: Apnea
    - No: Any period ($e_{AS} < r_{ap}$) > 10s?
      - Yes: Hypopnea
      - No: Normal breathing

Movement data
During preprocessing, any heart sounds and noise from the raw audio recordings are removed. To do so, a FIR bandpass filter with cutoff frequencies between 200 and 2000 Hz is applied to the raw audio signal. The filter limits have been chosen based on related studies found in literature [47, 50, 72]. Since breathing sounds range from loud to very shallow, background noise can easily interfere with the breath evaluation. Thus, a filtering technique called spectral subtraction is utilized to remove noise [9]. The resulting audio signal contains solely breathing sounds as exemplified in Figure 7.
Drop Detection

Figure 8: Detection of a possible apnea phase (PAS). The envelope $E_1$ (black) and envelope $E_2$ (grey) are calculated to detect drops of breathing amplitude. The dotted sections of $E_1$ indicate cut off values due to snoring suppression. The grey area marks the detected drop and its adjacent segments resulting in an extracted PAS [31, p. 5].

After preprocessing, the algorithm continues with the so called drop detection which is inspired by the procedures presented by Alshaer et al. [1, 2]. Here, potential apnea or hypopnea events are identified by scanning the entire signal for drops in breathing amplitude. In order to facilitate understanding, the key steps of the drop detection are exemplified in Figure 8. To capture short-term changes, like breathing cycles, the curve $E_1$ is computed by calculating the mean intensity of short-term windows. $E_1$ is represented by the black curve in Figure 8. Irregularities (e.g. snoring...
or artifacts) do not correlate to the actual amount of airflow when compared to normal breathing and are therefore cut-off. The black dotted curve in Figure 8 represents the corresponding cut-off values. In order to capture long-term changes in breathing and therefore flow, a second curve \( E^2 \) is computed. \( E^2 \) is calculated by interpolating the local maxima of single breathing cycles in the (truncated) first envelope \( E_1 \) to a curve using the Piecewise Cubic Hermite interpolation method. The envelope \( E^2 \) can now be used to identify drops in breathing. The detected drops and their adjacent segments are extracted as so-called possible apnea segment (PAS) in order to enable a more detailed examination.

The drop detection is especially designed to capture a very broad spectrum of drops in breathing. This might include drops in breathing that are neither apnea nor hypopnea but simply shallow breathing. Therefore, the final task of the algorithm is to correctly classify the detected PASs into apnea, hypopnea or normal breathing.
2 Materials and Methods

Classification

Figure 9: Event classification of a possible apnea segment (PAS). Curve $E3$ represents the audio power used to detect audio events. The grey areas mark the detected audio events and their heights represent estimated relative flow. The relative flow of the audio events in the reference segments ($RS$) define a threshold to classify the events in the apnea segment ($AS$) in breathing and non-breathing. In this example, all $AS$ events undershoot this threshold and are defined as non-breathing and therefore an apnea is detected. [31, p. 5].

The previously extracted PAS is now examined in detail in order to finally distinguish between apnea, hypopnea and normal breathing. This segment of the algorithm is referred to as classification. For better understanding, Figure 9 exemplifies the major steps of the classification algorithm. At first, every audio event $e$, which represents breathing, needs to
be identified. To do so, the power curve $E3$ is calculated and illustrated by the grey curve in Figure 9. Applying a threshold operation on $E3$ all audio events $e$ within the PAS are revealed. All audio events $e$ are represented by grey areas in Figure 9. The activity signal derived from the IMU is now incorporated in order to detect and remove audio events not caused by breathing but movement during the night.

To now detect apneas and hypopneas, their general definitions need to be applied (apneas, at least 90 %; hypopneas, at least 30 % reduction in air flow) to the body sound signal. In order to perform flow estimation using body sound, a reference value for normal breathing sound is essential. This is due to the fact that breathing sounds can change significantly over the course of one night (e.g. different sleeping positions) and especially between different individuals. Therefore, the previously detected PAS is split into segments to measure a reference for normal breathing (reference segment $RS$) and a segment to evaluate for apneas and hypopneas (apnea/hypopnea segment $AS$). An example for a split PAS is illustrated in Figure 9. If the overall duration of $AS$ is below 10 seconds, apneas or hypopneas are by definition not possible and the analysis is aborted. Otherwise all audio events $e$ are assigned to their corresponding segments and are labeled $e_{AS}$ and $e_{RS}$.

The algorithm continues by relating the tracheal body sound to actual airflow. Several studies have already investigated the relationship between airflow and breathing sounds and proposed several methods to correlate
the two signals [26, 68, 73]. However, these methods didn’t provide satisfactory results if applied to the tracheal body sound signal used here. Therefore, a new technique to calculate a feature value for airflow from audio signals has been developed and described in [31]. Now, a reference for normal breathing can be calculated by computing the median over the feature values of all events $e_{RS}$. Subsequently, the threshold for hypopneas $r_{hyp}$ (30 % reduction) and the threshold for apneas $r_{ap}$ (90 % reduction) are calculated. Finally, all events $e_{AS}$ can be classified into the categories, normal breathing, apnea event, and hypopnea event by comparing their feature value against the thresholds $r_{ap}$ and $r_{hyp}$. Any phase exceeding a duration of 10 seconds containing no normal breathing, and any hypopnea event, is marked as actual hypopnea whereas a phase containing no normal breathing and only apnea events is marked as actual apnea. After processing the entire recording over the course of one night the AHI can be calculated as apneas and hypopneas per hour of sleep.

2.2.2 Heart rate calculation

With the body sound microphone positioned at the neck, heart sounds can be recorded in addition to breathing sounds. These heart sounds can be used to measure the heart rate which is exceptionally important for the automated sleep staging described later. Figure 10 exemplifies the key
steps of the developed algorithm. Two distinct peaks can be recognized in the audio signal for every physiological heart beat. Usually these peaks are fairly small in amplitude compared to breathing or snoring. Therefore, the first step is to filter the raw audio signal using a bandpass filter with cutoff frequencies at 5 and 30 Hz. Filtering with such a narrow frequency band is necessary to remove breathing, snoring and artifacts from the original raw signal. Figure 10 (A) illustrates a raw signal whereas Figure 10 (B) illustrates the signal after filtering with suppressed breathing and preserved heart sounds.

To identify heartbeats, first all peaks with a defined minimal distance and an adaptive minimum height are detected. To assign two related peaks to a heartbeat, the distance from each peak to the adjacent peaks is evaluated. The two peaks with the minimal distance are then identified as one heart beat. Figure 10 (B) displays the detected peaks as stars, and a pair of peaks resulting in a heartbeat as grey circles. Even though, the presented method is robust, even during snoring, other artifacts such as bad microphone coupling or movements can have significant impact on the audio signal and cause the heart beat detection to fail. In those cases the heart rate is interpolated based on the last 10 preceding values until the artifacts cease.
Figure 10: Signal processing for pulse calculation. (A) raw audio signal; (B) audio signal after filtering in frequency domain, detected peaks are marked by stars, detected heart beat consisting of two peaks are marked by circles.
2.2.3 Movements and sleeping position

The IMU MPU-6000 (InvenSense) provides three gyroscope values and three accelerometer values with a sampling rate of 200 Hz. Since the IMU is placed on a thoracic belt at the chest, this data can be utilized to measure the sleeping position and the movement activity. To prevent common IMU interferences, like data drift, the methods presented by Madgwick et al. [42] have been established to receive stable results. Using these methods the orientation of the IMU and therefore of the sensor is provided in quaternion format.

The movement activity is described by the change of sensor orientation over a set amount of time. This activity scales from 0 (no movement) to 20 (great movement). Additionally, the sleeping position can be derived from the current sensor orientation. The number of changes in sleeping position over time are also determined for later sleep staging.
2.2.4 Sleep staging

In addition to the gold standard EEG, a large number of alternative signals and their characteristics can be used for sleep staging. [61]. Utilizing the developed ASM, methods for measuring respiration, heart rate, sleeping position, and movements have been presented in this work. It is suggested, that this information can be used to perform automated sleep staging in addition to AHI measurements. Existing research has proved that the dynamics of heart rate [10, 11, 52, 70] can be related to sleep stages. In addition, the dynamics of breathing as well as the respiratory frequency change over the course of a night depending on the sleep stages [20, 67]. Multiple studies have performed sleep staging based on this information and provide a huge collection of different cardiorespiratory features. To choose the optimal features from this huge collection, the feature selection methods presented by Khalighi S. et al. were utilized [34]. An overview of all used features and their corresponding sources can be found in Table 3.

The term NN-Interval is used in the following to describe the time difference between consecutive normal heartbeats. The time difference between successive breaths is referred to as the Breath-to-Breath interval (BB-Interval). The Methods presented in Chapter 2.2.2 are utilized to detect heart beats using the body sound signals provided. With this information the NN-Interval features presented in Table 3 can be calculated.
To calculate BB-Interval related and basic respiratory related features the techniques presented in Chapter 2.2.1 are used. Figure 11 illustrates signal processing from tracheal body sound signal to estimated airflow. With the estimated airflow signal, all corresponding features presented in Table 3 can be calculated. As for movement related features the IMU is utilized as described in Chapter 2.2.3. For a more detailed insight on feature calculation refer to the corresponding references.

The actual classification of the sleep stages is performed using a linear discriminant classifier (LDC). These classifiers have been used widely and successfully to perform automated sleep staging with similar features [18, 21, 39, 40, 59, 60]. The LDC is designed to assign a sleep stage to every 30 second epoch, like conventional sleep staging. For testing, as well as for training the LDC, the 30 features presented in Table 3 are calculated for each individual epoch. Additionally, three individual classifier systems for different sleep staging depth are designed. These include the 2-stage system which consists only of the stages wake (W) and sleep (REM, N1, N2, N3), the 3-stage system which consists of the stages wake (W), REM (REM) and NREM (N1, N2, N3) as well as the 4-stage system which consists of the stages wake (W), REM (REM), light sleep (N1, N2), deep sleep (N3).
Table 3: Features used for automated sleep staging; BB-Interval, interval between breaths; NN-Interval, interval between normal heartbeats; VLF, very low frequency; LF, low frequency; HF, high frequency; SD, standard deviation.

<table>
<thead>
<tr>
<th>Feature Category</th>
<th>Feature Number</th>
<th>Feature Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN-Interval frequency</td>
<td>1</td>
<td>spectral power of VLF (0.01–0.05 Hz)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>spectral power of LF (0.05–0.15 Hz)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>spectral power of HF (0.15–0.5 Hz)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>LF/HF-Ratio</td>
<td>[59, 60]</td>
</tr>
<tr>
<td>NN-Interval time features</td>
<td>5</td>
<td>NN-Interval mean</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>NN-Interval SD</td>
<td></td>
</tr>
<tr>
<td>BB-Interval frequency</td>
<td>7</td>
<td>spectral power of VLF (0.01–0.05 Hz)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>spectral power of LF (0.05–0.15 Hz)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>spectral power of HF (0.15–0.5 Hz)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>LF/HF-Ratio</td>
<td></td>
</tr>
<tr>
<td>BB-Interval time features</td>
<td>11</td>
<td>BB-Interval mean</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>BB-Interval SD</td>
<td></td>
</tr>
<tr>
<td>NN-BB cross spectral</td>
<td>13</td>
<td>cross spectral power of VLF</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>cross spectral power of LF</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>cross spectral power of HF</td>
<td></td>
</tr>
<tr>
<td>Respiratory</td>
<td>16</td>
<td>breath length mean</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>breath length SD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>breath by breath correlation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>breath by breath variation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>variation of the respiration frequency over 150</td>
<td>[59]</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>variation of the respiration frequency over 210</td>
<td></td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>variation of the respiration frequency over 300</td>
<td></td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>median estimated flow during breath cycles</td>
<td>[38]</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>median estimated flow during inhalations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>median estimated flow during exhalations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>ratio flow inhalation/exhalation</td>
<td></td>
</tr>
<tr>
<td>Movement</td>
<td>27</td>
<td>prominent sleep position</td>
<td>this work</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>number of sleep position changes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>mean acceleration across $x$, $y$, $z$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>mean angular velocity across $\Theta x$, $\Theta y$, $\Theta z$</td>
<td></td>
</tr>
</tbody>
</table>
Figure 11: Signal processing of the tracheal body sound for respiratory feature extraction. (A) raw audio signal, (B) audio signal after FIR-filtering and spectral subtraction, (C) estimation of airflow, values below the horizontal line are considered no breathing.
2.3 Clinical trial

In order to test the developed hardware and software solutions, a study based on the gold standard PSG was conceived and carried out. Therefore, the following sections give detailed information about the clinical trial which was performed. This includes the details of the trial registration, the study design, specification of the subject selection process, the description of measurement setup and the definition of the performed statistical evaluation.

2.3.1 Clinical trial registration

This study was approved by the ethics committee of the University of Ulm and all subjects gave written informed consent. The performed trial is also registered at the German Clinical Trials Register (DRKS):

- Trial name: Validation of a new method for ambulant diagnosis of sleep related breathing disorders using body sound.
- URL: https://www.drks.de/drks_web/navigate.do?navigationId=trial.HTML&TRIAL_ID=DRKS00011195
- Trial-ID: DRKS00011195
2 Materials and Methods

2.3.2 Study design

The study was designed as a monocentric, non-interventional, single arm investigator initiated trial (IIT) for basic research. The primary aim of this trial was to verify if the ASM was equivalent to PSG in the diagnosis of sleep-related respiratory disorders. The secondary aims were to verify if the ASM could be used reliably and without any complications and if there were other correlations between the data of the ASM and PSG.

2.3.3 Subjects

The study to validate the ASM and the developed algorithms was carried out at the sleep center of the University Hospital Ulm and included 60 subjects. All subjects included were referred to the sleep center with a suspicion of sleep related breathing disorders which had been diagnosed by a primary care physician. In accordance with the inclusion criteria all subjects were aged between 18 and 90 years. Subjects with known allergies or intolerances to adhesive patches, serious illness, or diseases that would affect participation or endanger the subject were excluded according to the exclusion criteria. Subjects included in the study underwent a standard full-night diagnostic PSG with the ASM setup simultaneously.

For validating the automated AHI estimation, four recordings were excluded. This was due to faulty body sound (n = 1), faulty airflow (n = 2)
or faulty thoracic and abdominal respiratory (n = 1) recordings. Patients suffering from central sleep apnea or mixed forms (n=4) and patients suffering from Cheyne–Stokes respiration (n=2) were also excluded. Of the remaining 50 subjects, 13 subjects suffered from mild, 11 from moderate, and 15 from severe OSA. The remaining 11 subjects were not diagnosed with OSA. More detailed anthropometric information of the included subjects is shown in Table 4.

For validating the automated sleep staging seven recordings were excluded. This was due to faulty body sound (n = 1) and faulty EEG (n = 6) recordings. Of the remaining 53 subjects, 14 subjects suffered from mild, 12 from moderate, and 16 from severe OSA. The remaining 11 subjects were not diagnosed with OSA. More detailed anthropometric information of the included subjects is shown in Table 5.
Table 4: Demographic information from the subjects included for validating the automated AHI estimation. Information shown here are based on the most common features used for obstructive sleep apnea (OSA) diagnosis according to [54]; BMI, body mass index; AHI, apnea hypopnea index; ET, evaluation time, contains only artifact-free periods, only this time is considered for calculation of characteristic values; TST, total sleep time; WASO, wake after sleep onset; N1, N2, N3, non-REM sleep stages; REM, rapid-eye movement sleep stage; ODI, oxygen desaturation index, number of desaturation events per hour; T90, time while SpO2 < 90 % [30, p. 1126].

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Gender(M/F)</td>
<td>31/19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>Age (years)</td>
<td>57.42</td>
<td>58</td>
<td>14.24</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>31.27</td>
<td>31.18</td>
<td>6.58</td>
</tr>
<tr>
<td>AHI</td>
<td>21.38</td>
<td>16.29</td>
<td>18.69</td>
</tr>
<tr>
<td>ET (hh:mm)</td>
<td>07:14</td>
<td>07:14</td>
<td>00:35</td>
</tr>
<tr>
<td>TST (hh:mm)</td>
<td>05:03</td>
<td>05:22</td>
<td>01:19</td>
</tr>
<tr>
<td>WASO (hh:mm)</td>
<td>01:48</td>
<td>01:20</td>
<td>01:08</td>
</tr>
<tr>
<td>Number of awakening events</td>
<td>33.2</td>
<td>30.5</td>
<td>14.58</td>
</tr>
<tr>
<td>Sleep efficiency (%)</td>
<td>69.89</td>
<td>76</td>
<td>17.08</td>
</tr>
<tr>
<td>N1 (%)</td>
<td>18.01</td>
<td>15.9</td>
<td>16.27</td>
</tr>
<tr>
<td>N2 (%)</td>
<td>55.41</td>
<td>57.15</td>
<td>16.49</td>
</tr>
<tr>
<td>N3 (%)</td>
<td>7.85</td>
<td>6.55</td>
<td>7.17</td>
</tr>
<tr>
<td>REM (%)</td>
<td>18.72</td>
<td>15.6</td>
<td>11.19</td>
</tr>
<tr>
<td>ODI (event/h)</td>
<td>19.03</td>
<td>12.05</td>
<td>19.29</td>
</tr>
<tr>
<td>T90 (hh:mm)</td>
<td>00:33</td>
<td>00:09</td>
<td>00:59</td>
</tr>
</tbody>
</table>
Table 5: Demographic information of the subjects included for validating the automated sleep staging; BMI, Body Mass Index; TST, total sleep time; TWT, total wake time; WASO, wake after sleep onset; SE, sleep efficiency; N1, N2, N3, non-REM sleep stages; REM, rapid-eye movement sleep stage.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender(M/F)</td>
<td>33/20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>57.3</td>
<td>57</td>
<td>14.07</td>
</tr>
<tr>
<td>BMI</td>
<td>31.24</td>
<td>31.14</td>
<td>6.39</td>
</tr>
<tr>
<td>TST (min)</td>
<td>299</td>
<td>317</td>
<td>80</td>
</tr>
<tr>
<td>TWT (min)</td>
<td>120</td>
<td>96</td>
<td>71</td>
</tr>
<tr>
<td>WASO (min)</td>
<td>112</td>
<td>91</td>
<td>70</td>
</tr>
<tr>
<td>SE (%)</td>
<td>71.21</td>
<td>77.98</td>
<td>16.56</td>
</tr>
<tr>
<td>N1/N2 (%)</td>
<td>50.12</td>
<td>52.07</td>
<td>13.04</td>
</tr>
<tr>
<td>N3 (%)</td>
<td>5.78</td>
<td>4.25</td>
<td>5.68</td>
</tr>
<tr>
<td>REM (%)</td>
<td>13.08</td>
<td>11.53</td>
<td>8.86</td>
</tr>
<tr>
<td>WAKE (%)</td>
<td>31.01</td>
<td>24.61</td>
<td>17.07</td>
</tr>
</tbody>
</table>
2 Materials and Methods

2.3.4 Data acquisition

The medical staff at the sleep center of the University Hospital Ulm were trained by technicians to perform the setup of the developed ASM independently. Both PSG and the ASM were set up by the trained medical staff and both recordings were monitored during the recording. Recordings were performed concurrently over the course of one night. The recordings were started between 9pm and 11pm and ended between 5am and 7am. The SOMNOlab system (Co. Weinmann Geräte für Medizin GmbH + Co. KG, Kronsaulswe 40, 22525 Hamburg, Germany) was utilized to perform PSG. This PSG included EEG, including channels C3-A2 and C4-A1 sampled at a rate of 256 Hz, unilateral anterior tibial EMG and bilateral EOG sampled at 256 Hz, oronasal airflow recorded using a thermistor and sampled at 32 Hz, thoracic and abdominal respiratory movements measured by using respiratory inductance plethysmograph and sampled at 32 Hz, oxygen saturation recorded by using finger pulse oximetry and sampled at 16 Hz as well as a basic heart rate monitoring using a 1-lead electrocardiograph sampled at 256 Hz. The ASM was attached to the thoracic belt included in the PSG setup. Tracheal body sound was recorded by the body sound microphone attached to the neck and sampled at 5 kHz. Actigraphy was carried out using the IMU and sampled at a rate of 250 Hz. Audio and movement data were transmitted wirelessly to a laptop for storage and subsequent data analysis.
Apnea and hypopnea scoring for PSG was performed in accordance with the AASM standards [4] which define apneas by \( \geq 90 \% \) drop of baseline flow amplitude with a duration of at least 10 seconds and hypopneas by \( \geq 30 \% \) drop of baseline flow amplitude with a duration of at least 10 seconds including an event related arousal and/or \( \geq 3 \% \) oxygen desaturation. A medical technician manually reviewed all data sets after the recording. The technician was blinded to the results of the ASM. After scoring, the AHI was automatically calculated as the number of apneas and hypopneas per hour of sleep. The sleep staging for PSG was performed by a trained technician who manually evaluated EEG, EOG and EMG recordings. Each 30 second epoch was assigned to a sleep stage in accordance with the AASM standards [4]. Human sleep was classified as the stages wake (W), rapid-eye movement (REM) and three non-REM stages (N1, N2, N3). The technicians were blinded to the results of the ASM. All ASM evaluations were performed with the previously described automated algorithms and independent of the results of PSG.

2.3.5 Statistics

MATLAB R-2015b (The MathWorks Inc., Natick, Massachusetts, USA) was utilized to perform the statistical analysis presented. For AHI analysis, the AHI measured by PSG is referred to as \( \text{AHI}_{\text{PSG}} \) and the AHI estimated by ASM is referred to as \( \text{AHI}_{\text{est}} \). To evaluate the agreement...
between $\text{AHI}_{PSG}$ and $\text{AHI}_{est}$ an equivalence test with equivalence limits of $\pm 5$ AHI at a significance level of 0.05 was performed. Additionally, Bland-Altman, as well as a correlation analysis, was carried out. To further investigate the overall performance of the ASM, the number of individually correctly and incorrectly classified apneas and hypopneas was calculated. Subjects were also classified into categories of mild ($5 < \text{AHI} > 15$), moderate ($15 < \text{AHI} > 30$) and severe ($\text{AHI} > 30$) OSA. For the evaluation of this classification, the sensitivity, specificity, positive predictive and negative predictive value, and the un-weighted Cohen’s kappa coefficient [37] were calculated. Additionally, the receiver operating characteristic (ROC) curves and the corresponding area under the curve (AUC) were calculated to evaluate the performance against the PSG results.

The validation of the LDC for automated sleep staging was carried out utilizing a leave one-out cross validation. This implies using 52 subjects for training and the remaining subject for testing the classifier and repeating the process until every subject has been used for testing once. After testing, the classification accuracy for every sleep stage can be calculated. The 2-stage, 3-stage and 4-stage sleep classifier systems were validated utilizing this process. In addition to accuracy, the Cohen’s kappa coefficient [37] can be computed for the classification of each individual sleep stage. This coefficient is a proper measure for unevenly distributed data, like sleep stages. Sleep stages are usually utilized to calculate several
parameters, like SE (ratio of sleep time to total time in bed). SE_{est} can be calculated utilizing the 2-stage system classifier and is evaluated by comparing it against PSG results SE_{PSG} using correlation analysis. The common sleep related parameters, wake after sleep onset (WASO), TST and total wake time (TWT) are also calculated and compared to the gold standard PSG. For use without medical supervision at home, SE groups are defined as an easy to use indicator. Based on the results of the 2-stage classifier subjects were grouped into categories of \( SE < 40 \% \), 40 \% < \( SE \) > 60 \%, 60 \% < \( SE \) > 80 \% and \( SE \) > 80 \%. These results were compared to the results of the PSG and the sensitivity, specificity, positive predictive value, negative predictive value and the un-weighted Cohen’s kappa coefficient were calculated. Additionally, the ROC curves and the corresponding AUCs were calculated to further evaluate the performance against the PSG results.
3 Results

3.1 Apnea–hypopnea index

The results of the automated AHI estimation are presented in the following sections. As stated previously in Section 2.3.2, 60 subjects underwent diagnostic PSG as reference standard in concurrence with the ASM. As described in Section 2.3.3, 50 subjects were included for the evaluation of the automated AHI estimation.

3.1.1 Equivalence

To evaluate equivalence between ASM and PSG, an equivalence test is performed. The paired differences $D$ are calculated as difference of estimated (EST) $\text{AHI}_{\text{est}}$ and $\text{AHI}_{\text{PSG}}$. A Kolmogorov-Smirnov test revealed that $D$ comes from a standard normal distribution. Testing $D$ against the previously defined equivalence limits $\pm 5$ AHI reveals:
3 Results

\[ H_0 : D = 5, \quad H_1 : D < 5, \quad p = 4.44e^{-22}, \quad t = -16.62 \]
\[ H_0 : D = 5, \quad H_1 : D > -5, \quad p = 3.50e^{-14}, \quad t = 10.32 \]

Since both null hypotheses can be rejected, \( D \) falls within the equivalence interval and equivalence can be claimed. Additionally, a standard t-test of \( D \) reveals: \( p = 2.8e^{-03} \), 95 \% CI = \([-1.9153 - 0.4236] \), standard deviation (SD) = 2.62.

3.1.2 Correlation and Bland-Altmann Analysis

A correlation analysis is performed and the correspondent scatterplot is shown in Figure 12 (A). A strong correlation between \( \text{AHI}_{est} \) and \( \text{AHI}_{PSG} \) can be observed and is reflected in the coefficient of determination \( r^2 \) of 0.9871. Likewise, the scatterplot for the Bland-Altman analysis is shown in Figure 12 (B). This analysis reveals, a mean ± 1.96 SD difference between \( \text{AHI}_{est} \) and \( \text{AHI}_{PSG} \) of 1.2 ± 5.14 and a coefficient of variation (CV) of 13 \%. A minor negative bias at low AHI (10-20) and a positive bias at high AHI (>30) can be observed.
Figure 12: (A) Correlation between AHI of the ASM (AHI_{est}) and the AHI of the PSG (AHI_{PSG}); AHI, apnea-hypopnea index; \( r^2 \), coefficient of determination; \( n \), number of data points. (B) Bland-Altman analysis showing variance between AHI_{est} and AHI_{PSG}; horizontal lines indicate the bias and the limit of agreement (±1.96 SD); CV, coefficient of variation [30, p. 1127]
3 Results

3.1.3 Sensitivity and Specificity

Table 6: Confusion Matrix of the apnea and hypopnea classification. The *no event* category refers to neither apnea nor hypopnea detected. PSG, polysomnography; ASM, ambulant sleep monitor.

<table>
<thead>
<tr>
<th></th>
<th>Apnea</th>
<th>Hypopnea</th>
<th>no event</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apnea</td>
<td>3785</td>
<td>94</td>
<td>292</td>
</tr>
<tr>
<td>Hypopnea</td>
<td>16</td>
<td>803</td>
<td>140</td>
</tr>
<tr>
<td>no event</td>
<td>472</td>
<td>247</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>sensitivity</th>
<th>specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Cohen’s kappa [CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHI ≥ 5</td>
<td>0.9091</td>
<td>0.9487</td>
<td>0.8333</td>
<td>0.9737</td>
<td>0.83 [0.65-1.02]</td>
</tr>
<tr>
<td>AHI ≥ 15</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0 [1.0-1.0]</td>
</tr>
<tr>
<td>AHI ≥ 30</td>
<td>1.0</td>
<td>0.9333</td>
<td>0.9722</td>
<td>1.0</td>
<td>0.95 [0.86-1.05]</td>
</tr>
</tbody>
</table>

In Table 6 the confusion matrix for the classification of apneas and hypopneas using PSG and ASM is shown. With a total of 50 recordings, 4273 apneas and 1144 hypopneas were diagnosed using PSG. Thereof, the ASM classified 3785 apneas and 803 hypopneas correctly (true positive). 386 apneas and 156 hypopneas were incorrectly classified (false positive).

Finally, the classification of subjects into groups of mild, moderate and severe OSA was evaluated. These three groups were defined by the thresholds AHI ≥ 5, AHI ≥ 15 and AHI > 30. Table 7 shows a detailed performance analysis of the presented classification. With a Cohen’s Kappa...
of $> 0.81$ for all groups, an almost perfect agreement can be considered [37]. Furthermore, ROC curves were created for each category and are shown in Figure 13. AUCs of 0.9627, 1.0 and 0.9962 were calculated for each category respectively. These results, show that the ASM has a high sensitivity and specificity for classifying subjects regarding the severity of OSA.
3 Results

Figure 13: Receiver operating characteristics (ROC) based on scaling groups mild, moderate and severe obstructive sleep apnea (OSA) [30, p. 1128].
3.2 Sleep staging

The results of the automated sleep staging are presented in the following sections. As stated previously in Section 2.3.2, 60 subjects underwent diagnostic PSG including EEG and EOG as reference standard in concurrence with the ASM. As described in Section 2.3.3, 53 subjects were included for the evaluation of the automated sleep staging algorithm.

3.2.1 Staging performance

Table 8: Classification performance of the ASM regarding sleep staging. REM, rapid-eye movement sleep stage; NREM, non-REM sleep stages; CI, confidence interval value.

<table>
<thead>
<tr>
<th>Classifier System</th>
<th>Accuracy sleep stages (%)</th>
<th>Accuracy total (%)</th>
<th>Cohen's kappa [CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wake Sleep</td>
<td>82.4</td>
<td>88.7</td>
<td>0.69 [0.68 - 0.70]</td>
</tr>
<tr>
<td>3-stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wake REM NREM</td>
<td>68.4 64.0 82.1</td>
<td>76.3</td>
<td>0.42 [0.42 - 0.43]</td>
</tr>
<tr>
<td>4-stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wake REM Light sleep Deep sleep</td>
<td>56.9 28.7 67.0 41.5</td>
<td>56.5</td>
<td>0.36 [0.36 - 0.37]</td>
</tr>
</tbody>
</table>

Detailed results of the system performance of the automated sleep staging algorithm presented here can be seen in Table 8. Results were calculated according the methods described in Section 2.3.5. The total accuracy decreases with increasing complexity of the staging systems. Re-
3 Results

ferring to the Cohen’s Kappa the 2-stage system provides a substantial agreement, the 3-stage system provides a moderate agreement, and the 4-stage system only provides a fair agreement [37].

3.2.2 Sleep related parameters

For the calculation of $\text{SE}_{est}$ the results of the 2-stage classifier were used. To compare these results to the $\text{SE}_{PSG}$ a correlation analysis was performed. The analysis revealed a a coefficient of determination $r^2$ of 0.78 with corresponding scatterplot shown in Figure 14. Performing a standard t-test of the paired differences (normal distribution) revealed $p = 0.008$, 95% $CI = [0.945, 0.98]$, $SD = 9.14$. Additionally, the AHI is calculated for each subject and the scatterplot of the paired differences between $\text{AHI}_{est}$ and $\text{AHI}_{PSG}$ and $\text{SE}_{est}$ and $\text{SE}_{PSG}$ is shown in Figure 15. Likewise, for the calculation of the parameters WASO, TWT and TST the results of the 2-stage classifier were used. Figure 16 shows the correspondent correlation scatterplots with a coefficient of determination $r^2$ of 0.81, 0.79 and 0.85 respectively.
### 3 Results

Table 9: System performance based on the evaluation of the subject classification into different groups of sleep efficiency (SE) including all 53 subjects; PPV, positive predictive value; NPV, negative predictive value; CI, confidence interval.

<table>
<thead>
<tr>
<th>SE $\geq$ 40</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Cohen's kappa [CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>SE $\geq$ 60</td>
<td>0.9608</td>
<td>1.000</td>
<td>10.000</td>
<td>0.5000</td>
<td>0.65 [0.17-1.12]</td>
</tr>
<tr>
<td>SE $\geq$ 80</td>
<td>0.8333</td>
<td>10.000</td>
<td>10.000</td>
<td>0.9535</td>
<td>0.89 [0.73-1.04]</td>
</tr>
</tbody>
</table>

Finally, the classification of subjects into groups of different SEs is evaluated. For the classification into groups of SE $\geq$ 40 %, SE $\geq$ 60 % and SE $\geq$ 80 % a detailed evaluation is shown in Table 9. Again referring to Cohen’s Kappa, a substantial agreement was reached for all groups [37]. In addition, ROC curves were created for each group and are shown in Figure 17.
Figure 14: Relationship between sleep efficiency (SE) of the new sleep monitor (SE\textsubscript{est}) and the sleep efficiency of the polysomnography (SE\textsubscript{PSG}); \( r^2 \), coefficient of determination; \( n \), number of data points.
Figure 15: Scatterplot of paired differences between the estimated (est) results and polysomnography (PSG) results of apnea-hypopnea index (AHI) and sleep efficiency (SE) including 53 subjects.
Figure 16: Relationship between wake after sleep onset (WASO), total wake time (TWT) and total sleep time (TST) of the new sleep monitor (EST) and the polysomnography (PSG) including 53 subjects; $r^2$, coefficient of determination.
3 Results

Figure 17: Receiver operating characteristics (ROC) of subject classification into groups of sleep efficiency (SE) (SE > 40 %, > 60 %, > 80 %); AUC, area under curve.
4 Discussion

In the following sections the results which were achieved will be critically analyzed. The AHI and OSA study results, as well as the automated sleep staging results, will be discussed in the context of current literature. Furthermore, the general study advantages and limitations will be illustrated. Finally, an outlook as well as a summary and conclusion are given.

4.1 AHI and OSA measurement

The automated AHI measurement of the presented ASM was demonstrated and its diagnostic capabilities were tested against standard PSG. The results of the proposed study revealed that the ASM accurately estimates AHI and reliably diagnoses OSA and its severity. To guarantee simple setups and high comfort the device consists only of a small device fixed by a chest belt and a microphone attached to the neck, which also
reduces the effect on sleep quality in comparison to existing ambulatory diagnostic systems.

To accomplish these results, a new algorithm to automatically measure AHI was developed and described. This method consists of three unique parts, preprocessing, drop detection, and classification to determine the AHI while utilizing only the tracheal body sound signal and movement data. The preprocessing step removes most heart sounds and noise through bandpass filtering from the raw signal. During drop detection, all segments of reduced breathing are identified. The final classification step performs a detailed review of those previously extracted segments and distinguishes between apnea, hypopnea and normal breathing. Additionally, movement data recorded by the IMU is utilized to detect and suppress artifacts caused by movements.

In contrast to the method presented, most commercially available ambulant sleep monitors utilize nasal pressure transducers, thermistors or thoracoabdominal movement belts to measure airflow. A combination of those methods is usually performed during PSG and are the gold standard for measuring breathing during the night. Ambulant sleep monitors however, rely solely on one method for the detection of apneas and hypopneas, which provides poor results [7]. Due to the independence of breathing route and abdominal/ thoracic breathing, it is suggested that tracheal body sound provides better results. However, one issue with body sound is the changing signal depending on individual anatomy [24].
To overcome this restriction, the methods presented individually and continuously recalculate the correlation between sound amplitude and airflow during apnea and hypopnea classification.

Processing all recordings including 4274 apneas and 1144 hypopneas, the proposed algorithm classified 3785 (89 %) apneas and 803 (70 %) hypopneas correctly and 386 (9 %) apneas and 156 (14 %) hypopneas incorrectly. Of all false positive apneas 24 % were incorrectly classified hypopneas and of all false positive hypopneas 10 % were incorrectly classified apneas. The remaining false positives were normal breathing, incorrectly classified as either apnea or hypopnea. These results demonstrate the main drawback of the automated algorithm which is to correctly distinguish between apneas and hypopneas and reduced but still normal breathing. It is suggested that this imprecision is due to the absence of an oxygen signal for the detection of oxygen desaturation. Apneas and hypopneas are not only defined by breathing reduction but also by oxygen desaturation. Without this information their detection and separation is most likely inaccurate. However, a study presented in [62] suggests that additional oximetry might not necessarily increase the accuracy of OSA diagnosis. Nevertheless, since oximetry is a standard in sleep monitoring it might be necessary for future health funding of the ASM. Additionally, the manual evaluation of the PSG recordings is open to the subjectivity of the technician. This leads to another imprecision which cannot be compensated by the automated algorithm.
4 Discussion

Compared to the gold standard PSG, the results presented prove that the ASM is able to diagnose OSA reliably. The mean 1.96 SD difference between $\text{AHI}_{\text{est}}$ and $\text{AHI}_{\text{PSG}}$ was $1.2 \pm 5.14$. The equivalence test between both methods revealed p-Values <0.001 for testing against the equivalence limits. Based on these results an equivalence between $\text{AHI}_{\text{est}}$ and $\text{AHI}_{\text{PSG}}$ can be claimed. Additionally, the correlation analysis revealed a strong correlation between $\text{AHI}_{\text{est}}$ and $\text{AHI}_{\text{PSG}}$ with a coefficient of determination $r^2$ of 0.987. Furthermore, classifying subjects into groups of different severities of OSA (mild, moderate, severe) using the ASM resulted in an almost perfect agreement with PSG. These results clearly outperform previously introduced approaches based on breathing sounds [2, 6, 22, 47]. The main limitation of those methods is suggested to be the method of recording body sound. The previously proposed body sound methods utilize recording devices which are mainly prototypes, and their microphones are not especially designed for body sound recording. In contrast, this work utilizes a highly sensitive body sound microphone designed for the long-term monitoring of breathing sounds. Therefore, the ASM is able to record an audio signal which represents breathing activity highly accurately and reliably and therefore improves the detection of apneas and hypopneas. A further issue of previous studies is the disturbance of the body sound signal by artifacts. Additionally, changes in sleeping position also have a significant effect on the acoustic features of body sound [48] and are not handled in those studies. These chal-
lenges limit previous approaches to the use of acoustic features, which are only minimally influenced by changes in sleeping position and artifacts. The ASM overcomes these issues by detecting movements as well as changes in sleeping position and by compensating them during signal processing. In summary, the ASM shows distinctive agreement with PSG concerning AHI and provides a strong diagnostic ability in evaluating the severity of OSA.

Type 4 sleep monitors, like the presented ASM, are simplified monitoring devices consisting of up to two measuring channels [16]. These monitors are mostly rejected for professional sleep diagnosis since they are not standardized and offer poor diagnostic performance [15, 64]. It is generally advised against the usage of a type 4 sleep monitor for the definite diagnosis of OSA [53]. In contrast, type 3 sleep monitors see a lot of use for ambulatory sleep monitoring, or prescreening. These devices utilize at least 4 channels, usually including ECG, airflow, effort, oximetry and provide good results for the diagnosis of OSA compared to PSG [15, 44].

The diagnostic performance of the presented ASM, however, is superior to type 4 sleep monitors and comparable with type 3 sleep monitors. Therefore, the ASM combines the advantages of both methods, resulting in a type 3 quality diagnostic accuracy while maintaining a simplified and comfortable type 4 monitor setup utilizing only a single lead. In addition, the diagnosis can be performed in a fully automated way. This feature saves time, compared to the manual evaluation, and might enable non-
sleep specialists, like family doctors, to carry out a simple OSA diagnosis. Finally, based on the results presented, it is suggested that the ASM is not only fit for diagnosis or prescreening but also for therapy control in already diagnosed patients. This is however not part of the current study and therefore more research is required to prove those suggestions.

4.2 Sleep staging

Table 10: Comparable sleep staging results found in literature.

<table>
<thead>
<tr>
<th>Classifier System</th>
<th>Source</th>
<th>Accuracy [%]</th>
<th>Cohen’s Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-stage system</td>
<td>Redmond et al. 2007 [59]</td>
<td>89.0</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>Mendez et al. 2010 [46]</td>
<td>79.3</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Devot et al. 2010 [18]</td>
<td>86.8</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>Willemen et al. 2010 [71]</td>
<td>92.0</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>this work</td>
<td>86.9</td>
<td>0.69</td>
</tr>
<tr>
<td>3-stage system</td>
<td>Long et al. 2014 [39]</td>
<td>76.2</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>Redmond et al. 2007 [59]</td>
<td>76.1</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>Mendez et al. 2009 [45]</td>
<td>79.4</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>Willemen et al. 2014 [71]</td>
<td>81.0</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>this work</td>
<td>76.3</td>
<td>0.42</td>
</tr>
<tr>
<td>4-stage system</td>
<td>Long et al. 2014 [39]</td>
<td>63.8</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>Willemen et al. 2014 [71]</td>
<td>69.0</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>Hedner et al. 2011 [25]</td>
<td>65.4</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>this work</td>
<td>56.5</td>
<td>0.36</td>
</tr>
</tbody>
</table>
4 Discussion

The application of the ASM for automated sleep staging was demonstrated, and its performance was validated by comparison against standard PSG sleep staging. Three types of sleep stage classifiers were implemented and validated. Here, it is important to note that a general performance limitation of sleep staging based on cardiorespiratory signals compared to EEG is suggested by current research [41]. Another source of imprecision is the manual sleep staging of the PSG which is open to the subjectivity of the technician. The results show a substantial agreement with the PSG for SE classification. For the actual sleep staging, Table 10 shows the results of similar approaches using cardiorespiratory features. Here, it is important to note that those approaches utilize well established methods or gold standards for the recording of cardiorespiratory signals (e.g. ECG and/or thermistor) compared to the body sound microphone presented here. The comparison indicates that the 2 and 3-stage system achieves comparable results. However, the 4-stage system is surpassed by the results of the literature. Nevertheless, it is important to keep in mind that the new method presented in this work only utilizes a type 4 sleep monitor where tracheal body sound is recorded with a single lead and movement data to extract the accumulation of cardiorespiratory features. Additionally, certain essential sleep parameters, like SE, can be calculated utilizing the 2-stage system. For a preliminary screening, or a simple evaluation of sleep quality, it is suggested that the 2 or 3-stage system is sufficient.
4 Discussion

In recent years, some research on unobtrusive and comfortable cardiorespiratory sleep staging using alternative methods has been performed. A high-resolution pressure-sensitive bed sheet is presented by [65]. This bed sheet is utilized to extract sleep-related biophysical and geometric features for sleep staging. They achieved an overall accuracy of 71.1% for a 3-stage system, with seven subjects included in the study. Another researcher group presented a method which utilized sensor foils placed into the bed mattress [36]. They reached an accuracy of 79% and a Kappa of 0.44 for a 3-stage system, with 18 subjects included in their study. The results of those methods are similar compared to the results of the proposed ASM. However, these methods also come with some drawbacks, like excessive noise problems during body movements. Finally, the significance of these studies is limited by the low number of participants.

Some research uses subject-specific classifiers or subject-specific feature normalization to improve their automated sleep staging [39, 60]. Nevertheless, a subject-independent classifier (like the presented LDC) comes with the advantages of a setup without the need for calibration or any adjustments, again facilitating the use in the homecare area without medical supervision. For studies over the course of multiple nights however, subject-specific feature normalization could be useful. It is suggested that further research should investigate the advantages of subject-specific classifiers or subject-specific feature normalization.
4 Discussion

4.3 General study advantages and limitations

The study which was conducted has several advantages, but also limitations. The setup of both the ASM and the PSG was performed by previously trained medical staff. During the night, both recordings were performed simultaneously. Of all 60 recordings, four had to be discarded since an extensive amount of data had been lost or was faulty. Of those unusable recordings, only a single one was caused by the failure of the ASM. The remaining interferences were caused due to the failure of the recording channels of the PSG. This reinforces the advantages of the simple and reliable setup as well as stand-alone operation of the ASM. It is therefore suggested that the ASM can also be operated by subjects without extensive training and medical knowledge in a home setting. However, a future study in a home setting will be necessary to fully validate this suggestion.

The study only contains so called first diagnostic nights, in which all subjects visited the sleep center for the first time and PSG was performed without any additional therapeutic measures. The disadvantages of an unfamiliar environment, which have already been discussed, paired with the extensive recording equipment are suggested to be the cause of the high WASO and also the unusual low TST compared to normal healthy sleep. An additional note of caution has to be given, since only a thermistor is utilized by the PSG for airflow measurement. Some researchers
suggest that this can cause an underestimation of AHI [58]. It is therefore highly recommended to include a nasal transducer in future studies for a more accurate measurement of AHI.

Subjects in the study were only recruited if they had a suspicion of any form of sleep apnea and all recordings were performed in a clinical environment. The results are therefore limited to be only applicable to similar populations and may vary if applied to a more general population. Since one future application of the ASM might be the use for home screening, the subjects might not be an appropriate representation of the target group. However, it needs to be noted that the PSG results cover the entire spectrum, from healthy subjects to subjects suffering from severe OSA. In addition, age, sex and body mass index (BMI) distribution cover a broad spectrum of different individuals.

The ASM provides stable results for all participants for AHI measurements. For the performed sleep staging it needs to be noted that the sleep of subjects suffering from OSA is disrupted by arousals caused by breathing pauses, and therefore does not represent healthy sleep. However, a correlation analysis revealed no relationship between OSA and sleep staging performance. This might imply that the sleep staging performed with the ASM performs just as well for healthy subjects as for subjects suffering from OSA. Additionally, Redmond et al. [59, 60] also agrees with the suggestion that sleep staging based on cardiorespiratory signals can also be performed on patients suffering from OSA. However,
further studies are required to fully validate the proposed methods in a more diverse subject population, and in a home setting.

The study conducted was limited to patients with suspected OSA, therefore it was not possible to evaluate the diagnostic performance of the ASM in respect to other sleep related breathing disorders. Using sleep staging, however, it is suggested that the ASM can also be used to diagnose other sleep disorders (e.g. insomnia), or to perform a preliminary screening to decide whether a PSG is necessary or not. However, further studies need to be carried out to validate the diagnostic performance for those disorders. For OSA diagnosis, it is of immense importance to use only the ASM in subjects with a high OSA pretest probability. This should reduce the number of false-positive diagnoses and avoid false-negative diagnoses for patients suffering from other sleep related breathing disorders who are better suited for PSG.

4.4 Outlook

As already mentioned, one drawback of the ASM could be the missing oximetry. In the current state medical funding for a portable sleep monitor is only possible if oximetry is included. Adding another channel to the ASM is contrary to the simple nature of this device, but might be necessary for future development. A solution for this issue could be to integrate
the oximetry measurement into the cone of the body sound microphone. This would keep the ASM and its setup simple while adding another channel. The less commonly used reflectance pulse oximetry does not require a thin body part compared to the commonly known transmissive pulse oximetry which generally measures at the fingertip or earlobe. Current experimental investigations with this technology are being carried out to evaluate the possibility to measure oxygen at the neck for future integration in the ASM.

A preliminary manual inspection of the recorded IMU data revealed that ventilation movements can be captured utilizing the acceleration and velocity information. This is due to the positioning of the IMU at the upper respiratory inductance plethysmography band of the PSG during the study. Additionally, vibrations caused by snoring were also captured. It is therefore likely that the additional information of the IMU data can be utilized to distinguish between central sleep apnea and OSA, and to detect snoring. These unexpected outcomes were revealed after the conduct of the study and are currently not part of the proposed method. Therefore, more research is necessary to examine those suggestions.
4.5 Conclusion

OSA is the most prevalent form of the most common sleep disorder sleep apnea. The PSG is the gold standard for diagnosis but involves an overnight stay in a sleep laboratory causing high costs and providing limited sleep capacities, resulting in long waiting periods for patients worldwide. With an alarmingly high number of undiagnosed and untreated patients, paired with extensive and costly diagnosis, significant effort has been made to develop more simple, but still reliable, sleep diagnostic systems [56]. While ambulant sleep monitoring is possible, those applications are constrained by reduced diagnosis abilities or extensive setup, and cannot be operated without medical assistance and therefore greatly limit use in home areas.

To overcome those problems a new comfortable and simple sleep monitoring system for the automated diagnosis of OSA and basic sleep staging utilizing only tracheal body sounds and movement data is proposed and validated using PSG as gold standard. In conclusion, the study presented provides evidence that the proposed ASM can accurately and automatically calculate AHI, diagnose the presence and severity of OSA, and perform a basic sleep staging. This minimalistic approach can address the need for a simple, mobile and also reliable prescreening or diagnosis of OSA. However, its application in a home setting, one of the most important fields of application for the proposed monitor, has not yet
4 Discussion

been validated. Therefore, future studies need to be carried out to ensure reproducibility and applicability in an ambulant setting.
5 Abstract

The gold standard for assessment of most sleep disorders is the in-laboratory polysomnography (PSG). This approach produces high costs and inconveniences for patients due to its extensive setup, whereas alternative ambulatory systems are limited through reduced diagnostic abilities. The work presented here, therefore, aims to develop and validate a new, reliable, and simplified ambulant sleep monitor, utilizing tracheal body sound and movement data to automatically diagnose obstructive sleep apnea (OSA), one of the most common sleep disorders. To further improve the diagnostic ability of this monitor, automated sleep staging should be performed by utilizing body sound to extract cardiorespiratory features and actigraphy to extract movement features.

The main criteria to indicate the severity of OSA is the apnea-hypopnea index (AHI). Therefore, a new algorithm for the automated calculation of AHI was developed. For validation, the data of 60 subjects was recorded at the University Hospital Ulm. Subjects underwent a full-night screen-
Abstract

ing using PSG and the new monitoring system concurrently. The AHI was scored blindly by a medical technician using PSG ($\text{AHI}_{PSG}$) and by the automated algorithm ($\text{AHI}_{est}$). $\text{AHI}_{est}$ strongly correlates with $\text{AHI}_{PSG}$ ($r^2=0.9871$). A mean $\pm$ 1.96 SD difference between $\text{AHI}_{est}$ and $\text{AHI}_{PSG}$ of $1.2 \pm 5.14$ is achieved. In terms of classifying subjects into groups of mild, moderate and severe sleep apnea, the evaluated new sleep monitor shows a strong correlation with the results obtained by PSG (Cohen’s Kappa > 0.81). These results clearly outperform similar approaches which were previously used. Additionally, a linear discriminant classifier was used to perform automated sleep staging using the new sleep monitor. The classifier achieved 86.9% accuracy with a Kappa of 0.69 for sleep/wake classification, 76.3% accuracy with a Kappa of 0.42 for wake/REM/NREM classification and 56.5% accuracy with a Kappa of 0.36 for wake/REM/light sleep/deep sleep classification. For the calculation of sleep efficiency (SE) a coefficient of determination $r^2$ of 0.78 was reached. Here, subjects were also classified into groups of SEs ($\text{SE} \geq 40\%$, $\text{SE} \geq 60\%$ and $\text{SE} \geq 80\%$). A Cohen’s Kappa > 0.61 was achieved for all groups.

The proposed sleep monitor accurately estimates AHI and diagnoses sleep apnea and its severity reliably. Furthermore, the monitor provides good performance in sleep/wake and wake/REM/NREM sleep staging while maintaining a simple setup and offering high comfort.
6 Bibliography


[10] M. H. Bonnet and D. L. Arand, “Heart rate variability: Sleep stage,


[51] T. Penzel, G. Hajak, R. Hoffmann, R. Lund, T. Podszus,


[68] K. A. Sohrabi, D. Basu, F. Schudt, M. Scholtes, O. Seifert,


[73] A. Yadollahi, A. Azarbarzin, A. Montazeri, and Z. Moussavi, “Acous-


Acknowledgements

This section has been removed for privacy reasons.
Curriculum vitae

This section has been removed for privacy reasons.
This section has been removed for privacy reasons.
This section has been removed for privacy reasons.